ESSAYS ON NONPARAMETRIC IDENTIFICATION AND ESTIMATION OF CONTRACTS

A Dissertation

by

DAIQIANG ZHANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Qi Li
Co-Chair of Committee,	Yonghong An
Committee Members,	Fernando Luco
	Silvana Krasteva
	Ximing Wu
Head of Department,	Timothy Gronberg

May 2017

Major Subject: Economics

ABSTRACT

Due to the fundamental role of asymmetric information in economic relations, during the past decades contracts have flourished and dedicated to how the information asymmetry and incentives induce strategic behavior among economic agents in a number of directions. One branch of contracts focuses on bilateral contract between one principal and one agent, in which the agent's hidden information and hidden action lead to the adverse selection and moral hazard problems, respectively. There is a simple and widely used menu of contracts, which consists of two types of contracts: the fixed-price contract in which the payment is fixed regardless of the realized cost, and the cost-reimbursement contract in which the payment equals the realized cost.

By allowing for a more reasonable property that the optimal effort is monotone in the agent's type, we show that the performance of the optimal FPCR contract relies crucially on the cost function: when the marginal is relatively large, the gain of the optimal FPCR menu is very close to the fully optimal contract. Otherwise the optimal FPCR menu behaves arbitrarily close to a cost-reimbursement contract.

To quantify the renegotiation cost without commitment in a two-period setting, we nonparametrically identify and estimate the model primitives, including the agent's cost function and disutility function of effort, the distribution of innate, agents' bargaining power and the intertemporal preference. The empirical evidence shows that the nonlinearity of agents' cost function implies different empirical results about the distribution of welfare gains between firms and taxpayers.

The second branch of bilateral contracts between one principal and multiple agents emphasizes the externalities generated by the dependence of one agent's payoff on other agents' contracts, while the first branch of contracts involves no externalities due to the set-

ting of one principal and one agent. To quantify the payment effect of bargaining power, I nonparametrically identify and estimate the model primitives, including the manufacturer's cost function, the hospitals' payoff function, the joint distribution of hospitals' payoff-shocks, and hospitals' bargaining power. And, we conduct counterfactual analysis of the effect of the bargaining power on the price.

DEDICATION

To my mother Caifang Yan, my father Weifu Zhang, my wife Xue Gong, my two daughters Emma Zhang and Sophia Zhang.

ACKNOWLEDGMENTS

I am deeply indebted to Qi Li and Yonghong An for their guidance and encouragement on this project. I benefited greatly from the comments of Fernando Luco, Silvana Krasteva, and Vikram Manjunath. I thank Xavier d'Haultfoeuille, Kate Ho, Brad Larsen, Rosa Matzkin, Byoung Park, Geert Ridder, Xun Tang, and Yuanyuan Wan for their helpful comments.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Qi Li of the Department of Economics (chair), Professor Yonghong An of the Department of Economics (co-chair), Professor Fernando Luco of the Department of Economics, Professor Silvana Krasteva of the Department of Economics, and Professor Ximing Wu of the Department of Agricultural Economics.

The analyses depicted in Chapter one and Chapter two were conducted in part by Professor Yonghong An of the Department of Economics.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by the Department of Economics of Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Due to the fundamental role of asymmetric information in economic relations, during the past decades contracts have flourished and dedicated to how the information asymmetry and incentives induce strategic behavior among economic agents in a number of directions. Two of widely used contracts are fixed-price-cost-reimbursement (FPCR) contracts and one-principal-multiple-agents contracts with externalities.

The FPCR contract is a bilateral contract between one principal and one agent, in which the agent's hidden information and hidden action lead to the adverse selection and moral hazard problems, respectively. There is a simple and widely used menu of contracts, which consists of two types of contracts: the fixed-price contract in which the payment is fixed regardless of the realized cost, and the cost-reimbursement contract in which the payment equals the realized cost. The seminal paper [1] demonstrates that when the agent's innate cost is uniformly distributed, a simple menu of contract (FPCR menu) captures at least three-fourths of the surplus that a fully optimal contract proposed in [2] achieves. Nevertheless, the powerful conclusion of [1] is based on the assumption of agents' cost function being an identity, as also imposed in many other studies of contracts, e.g., [2] and [3]. While such a restriction simplifies the theoretical analysis, more realistic cost functions are preferable and have been considered in the contract theory. The influential work [4] assumes a strictly convex cost function, and more recently [5] adopt the same setting of cost function. Moreover, the constant effort level for agents with different innate costs is inconsistent with some empirical evidence. For example, [6] estimate the firms' effort to be increasing in their innate cost using the urban transport data in France. By allowing for a more reasonable property that the optimal effort is monotone in the agent's type, we show that the performance of the optimal FPCR contract relies crucially on the cost function: when the marginal is relatively large, the gain of the optimal FPCR menu is very close to the fully optimal contract. Otherwise the optimal FPCR menu behaves arbitrarily close to a cost-reimbursement contract.

Besides, in reality the FPCR contracts are sign for more than two periods, and there exists renegotiation of contract among the multiple periods. In theory, the renegotiation leads to loss of social welfare, and the magnitude of this loss is important for policy implications. To quantify the renegotiation cost without commitment in a two-period setting, we nonparametrically identify and estimate the model primitives, including the agent's cost function and disutility function of effort, the distribution of innate, agents' bargaining power and the intertemporal preference. The empirical evidence shows that the nonlinearity of agents' cost function implies different empirical results about the distribution of welfare gains between firms and taxpayers.

The second branch of bilateral contracts between one principal and multiple agents emphasizes the externalities generated by the dependence of one agent's payoff on other agents' contracts, while the first branch of contracts involves no externalities due to the setting of one principal and one agent. We consider the contracting with externalities developed by [7]. In the model, the principal simultaneously makes one offer to each agent, and then agents simultaneously decide whether to accept or reject their own offers. Because those offers are private in the sense that each agent only observes his own offer, it is a dynamic incomplete-information game. In actuality, the model has found wide applications in various economic situations, including vertical contracts in which profits of downstream firms depend on all downstream firms' contracts with the upstream firm ([8]; [9]); exclusive dealing in which payoffs of agents rely on the number of agents who sign exclusive contracts with the principal ([10]; [11]); network externalities ([12]; [13]); and among others. Despite the wide use of contracting with externalities in various sectors such as medical device industry and publishing industry, there are few empirical studies on the incomplete-information contracting game with externalities. This paper constitutes the first effort on the rigorous econometric analysis of contracting with externalities. To quantify the payment effect of bargaining power, I nonparametrically identify and estimate the model primitives, including the manufacturer's cost function, the hospitals' payoff function, the joint distribution of hospitals' payoff-shocks, and hospitals' bargaining power. And, we conduct counterfactual analysis of the effect of the bargaining power on the price. By applying the model to coronary stents contracts between the manufacturer and hospitals in the United States, we find that passive beliefs fit the dataset better than symmetric beliefs. More relevantly, the counterfactual result shows that if the bargaining power of hospitals increases, under passive beliefs the stent's price decreases by some reasonable amount, while under symmetric beliefs the decrease of price is unreasonable.

2. SIMPLE MENUS OF COST-BASED CONTRACTS WITH CONVEX COST FUNCTIONS

2.1 Introduction

Since the influential work [2] on the principal-agent model of procurement and regulation, a vast literature has been devoted to studying the performance of simple mechanisms in Laffont-Tirole type principal-agent models. The seminal paper [1] demonstrates that when the agent's innate cost is uniformly distributed, a simple menu of contract (FPCR menu) captures at least three-fourths of the surplus that a fully optimal contract achieves. The FPCR menu consists of two simple contracts: a fixed-price (FP) contract where the payment to the agent is a fixed price, regardless of the agent's realized cost, and a cost-reimbursement (CR) contract where the agent is reimbursed exactly the realized cost.

The simple contracts in [1] are of great practical importance because the fully optimal contract proposed in [2] is too complex to be implemented in practice, whereas the aforementioned much simpler menu at least in some cases, secures a substantial share of the surplus that a fully optimal menu can secure. Nevertheless, the powerful conclusion of [1] is based on the assumption of agents' cost function being an identity, as also imposed in many other studies of contracts, e.g., [2] and [3]. While such a restriction simplifies the theoretical analysis, more realistic cost functions are preferable and have been considered in the contract theory. The influential work [4] assumes a strictly convex cost function, and more recently [5] adopt the same setting of cost function. Moreover, the constant effort level for agents with different innate costs is inconsistent with some empirical evidence. For example, [6] estimate the firms' effort to be increasing in their innate cost using the urban transport data in France.

Motivated by both the existing theoretical work and the empirical evidence, we relax the assumption of identity cost function in the model of FPCR menu and investigate how the performance of the optimal FPCR menu would be impacted by a convex cost function. We provide an important observation for the optimal FPCR menu when the agent's innate cost is uniformly distributed: the performance of the optimal FPCR contract relies crucially on the cost function: when the marginal cost (relative to marginal disutility) is large, the performance is very close to that of the fully optimal contract. On the other hand, if the marginal cost is small and firms' innate costs are dispersed then the performance of the optimal FPCR menu is arbitrarily close to a CR contract. Our result is in contrast with that of [1], where the FPCR menu captures at least three-fourths of the surplus that a fully optimal contract achieves. The main force that leads to the discrepancy is that under a convex cost function, the optimal cost-reducing effort exerted by an agent is strictly increasing in her innate cost and a cost function with higher marginal cost induces larger cost-reducing effort for given innate cost. By contrast, an identity cost function implies that the optimal cost-reducing effort is a constant, regardless of the innate cost of the agent. Our finding suggests that in designing an optimal FPCR contract it is important for the principal to take into account the cost structure of the agents: when the marginal cost of agents is large, the FPCR menu is especially preferable. When the marginal cost is small and agents' innate costs are dispersed, the menu is less appealing or even arbitrarily close to a CR contract.

The remaining of this paper is organized as follows. Section2 presents the basic model and some of its properties. Section 3 compares the performance of the optimal FPCR menu with the fully optimal contract. Section 4 concludes and proofs are included in the Appendix.

2.2 The Model

A principal wishes to procure a project by offering a menu of contracts to a firm (agent).[1] The firm's realized cost of the project is $c = H(\theta - e)$, where θ and e are the agent's innate cost and cost-reducing effort, respectively, and $\theta - e > 0$. Exerting effort e incurs disutility $\psi(e)$, thus the total cost of the agent for the project is $H(\theta - e) + \psi(e)$. Let $t(c)$ be the payment from the principal to the agent in a FPCR menu, then $t(c) = b$ for FP contracts and $t(c) = c$ for CR contracts, where b is the fixed-price and c is the realized cost. A firm who accepts the CR contract only gets the realized cost c reimbursed, hence this firm exerts no effort and makes zero profit, whereas a firm who chooses the FP contract makes profit $\pi \equiv b - H(\theta - e) - \psi(e)$.

Assumption 1. *(i) Both cost and disutility functions are quadratic. Specifically, $H(\theta) = \beta\theta^2, \beta > 0$ and $\psi(\cdot)$ is normalized to be $\psi(e) = e^2$. (ii) A firm's innate cost θ is uniformly distributed on its support $[\underline{\theta}, \bar{\theta}] \subset \boldsymbol{R}_+$.*

Both $H(\cdot)$ and $\psi(\cdot)$ are standard cost functions in economics (e.g., [4] and [3]). Note that the convex cost $\beta(\theta - e)^2$ if effort is exerted also implies that the effort $e \geq 0$, otherwise, exerting effort increases the firm's cost. Under the CR contract, the firm's cost is reduced to $c = H(\theta)$ and the profit $\pi = c - c - \psi(0)$ must be zero for any realized cost c by the noting the assumption that $\psi(0) = 0$. The assumption of uniform distribution for firms' innate costs is imposed for comparison with [1].

Define e^* as the optimal positive cost-reducing effort exerted by a firm who accepts the FP contract. By employing the firm's first-order condition $H'(\theta - e^*) = \psi'(e^*)$, i.e., the marginal cost reduced by the effort equals the marginal disutility of exerting the effort, we obtain the optimal cost-reducing effort $e^*(\theta) = \beta\theta/(\beta + 1)$ with the strict monotonicity of cost-reducing effort $0 < e^{*'} < 1$. On the one hand, the optimal FPCR menu induces

[1] We use “firm" and “agent" exchangeably.

less efficient firms (with a larger θ) to exert more cost-reducing effort. In existing studies on the FPCR menu, however, the optimal effort e^* is a constant. This is because the cost $H(\cdot)$ is an identity function (e.g., see [1] and [14]), thus the first order condition of the firm $\psi(e^*) = 1$ is independent from θ. Intuitively, an identity cost function implies that the marginal benefit of exerting the same effort is the same for firms with different innate costs. Therefore, a firm has no incentive to exert more effort than others. However, when the cost is convex a firm with a higher innate cost enjoys a larger reduction of cost than a lower-innate cost firm by exerting the same effort, and this explains the strict monotonicity of cost-reducing effort in firms' innate cost. On the other hand, for a given innate cost θ, the larger β the higher level of optimal effort. This result can be explained as follows. A larger β implies higher marginal benefit from exerting the same effort, i.e., the marginal benefit $2\beta(\theta - e)$ is increasing in β, whereas the marginal disutility $2e$ does not depend on β. Therefore, the firm will exert more effort to minimize the cost if β is larger.

We now turn to the problem of the principal. Suppose that the principal offers the fixed-price b such that the cut-off type is θ^*, i.e., the firm with $\theta^* \in [\underline{\theta}, \bar{\theta}]$ is indifferent between choosing the FP and the CR contract. The profit of the firm with type $\theta = \theta^*$ $\pi = t(c) - H(\theta^* - e^*(\theta^*)) - \psi(e^*(\theta^*))$ is zero. Considering that $d\pi/d\theta = -2\beta\theta/(1+\beta) < 0$, all the firms with type $\theta \leq \theta^*$ choose the FP contract and make positive profit; whereas those with $\theta > \theta^*$ choose the CR contract and earn zero profit. For the FPCR menu with θ^* being the cut-off type, the principal's expected cost for the project is:

$$C(\theta) \quad = \quad \int_{\underline{\theta}}^{\theta} \left[H(\theta - e^*(\theta)) + \psi(e^*(\theta))\right] dF(v) + \int_{\theta}^{\bar{\theta}} H(v) dF(v), \tag{2.1}$$

where $F(v) = (v - \underline{\theta})/(\bar{\theta} - \underline{\theta})$ is the cdf of firms' type θ. The two terms on the right-hand-side of the second equality are the payment to firms accepting the FP and the CR contract, respectively.

Define $\gamma \equiv \bar{\theta}/\underline{\theta} \in (1, \infty)$, then a larger γ implies a higher average innate cost and more dispersed distribution. The optimal FPCR is stated as following.

Lemma 1. *Suppose Assumption 1 hold, in an optimal FPCR contract, the principal offers a fixed-price such that: (i). If $0 < \beta < 2(\gamma - 1)/\gamma$, all the firms with $\theta \leq \theta^* = 2\underline{\theta}/(2 - \beta)$ choose the FP contract and the remaining firms choose the CR contract. The fixed-price is $b = \frac{4\beta\underline{\theta}^2}{(2-\beta)^2(\beta+1)}$. (ii). If $2(\gamma - 1)/\gamma \leq \beta$, then $\theta^* = \bar{\theta}$. All the firms choose FP contract, and the fixed-price is $b = \frac{\beta}{\beta+1}\bar{\theta}^2$.*

Lemma 1 states that when β is small it is optimal for the principal to offer a FPCR menu that only induces some firms (those with lower innate costs) to choose the FP contract. When β is large, the optimal menu is such that all the firms choose the FP contract.

The intuition behind Lemma 1 is provided in the discussion after Assumption 1: a larger β means a higher marginal benefit from exerting the same effort, and this provides incentives for inefficient firms to exert more effort to reduce the cost. The optimal FPCR menu in Lemma 1 is a response to such impacts of β.

2.3 Comparison of Performance

In this section, we evaluate the performance of the optimal FPCR menu by comparing its surplus with the fully optimal contract (Laffont and Tirole, 1986, 1993) in a similar way to [1] and [14]. First, we conduct an analysis of the fully optimal contract with a convex cost function. [2, 4] characterize the fully optimal mechanism that can be implemented by the principal offering a menu consisting of a continuum of linear contracts. We follow the standard practice in their seminal work to obtain the cost-reducing effort $e^*(\theta)$ in Lemma 2.

Lemma 2. *Under Assumption 1, the optimal level of cost-reducing effort $e^*(\theta)$ in the fully* *optimal contract is:*

$$e^*(\theta) = \begin{cases} 0, \textit{if } \ \theta \geq \underline{\theta}/(1-\beta) \ \textit{ and } \ \beta \leq 1. \\ \frac{(\beta-1)\theta+\underline{\theta}}{\beta+1}, \textit{if } \ \theta \leq \underline{\theta}/(1-\beta) \ \textit{ and } \ \beta < 1, \ \textit{ or } \ \beta \geq 1. \end{cases} \tag{2.2}$$

Lemma 2 states that under the fully optimal contract there also exists a cut-off type such that only those more efficient types than the cut-off one exert effort. This result is similar to that of the optimal FPCR contract in Lemma 1. Specifically, only some firms exert effort when β is small, but when β is large all the firms minimize cost by exerting positive effort. Nevertheless, a comparison between two lemmas demonstrates that the cut-off types induced by the two contracts are different. Moreover, from Lemma 2 we have $e^{*\prime}(\theta) = (\beta - 1)/(\beta + 1) \leq 0$ when $\beta \leq 1$. This is consistent with the conclusion in [2] but opposite to the optimal FPCR menu. For the case of $\beta > 1$ both the fully optimal contract and the optimal FPCR one induce strictly increasing effort in firms' innate cost.

Let $G_F \equiv C(\underline{\theta}) - C(\theta^*)$ denote the reduction in the principal's expected cost under the optimal FPCR contract relative to the CR contract, and G_O is defined similarly for the fully optimal contract. We focus on the analysis of the ratio G_F/G_O, which describes the performance of the optimal FPCR menu relative to the fully optimal one.

Theorem 1. *For any given $\gamma > 1$, (i). If $0 < \beta \leq (\gamma-1)/\gamma$, G_F/G_O is strictly decreasing in β and $G_F/G_O < 3/4$. (ii). If $(\gamma - 1)/\gamma < \beta < 2(\gamma - 1)/\gamma$, G_F/G_O decreases to its minimum then increase in β. (iii) If $\beta \geq 2(\gamma - 1)/\gamma$, G_F/G_O is strictly increasing in β and $\lim_{\beta\to\infty} G_F/G_O = 1$.*

As expected, the performance of the FPCR contract depends on both the distribution parameter γ and the cost parameter β. To visualize the findings in Theorem 1, we illustrate the relationship between G_F/G_O and β for $\gamma = 3$ in panel (a) of Figure 2.1. The function consists of three segments: for $\beta \in (0, (\gamma-1)/\gamma]$, the ratio G_F/G_O strictly decreases in β, starting from $3/4$; in the second segment, $\beta \in ((\gamma-1)/\gamma, 2(\gamma-1)/\gamma)$, G_F/G_O declines first to reach the minimum, and then rises again. When $\beta \geq 2(\gamma-1)/\gamma$, G_F/G_O increases to one as β is getting large. In summary, as β increases the performance of the optimal FPCR contract is getting worse first, then reaching the lowest level, and finally catching up the fully optimal contract. Especially, when β is large, the optimal FPCR contract secures almost all the gain secured under the fully optimal contract.

To better understand the results in Theorem 1, we explore the economic reasonings of the patterns shown in pane (a) of Figure 2.1. The declining segment of G_F/G_O for $\beta \in (0, (\gamma-1)/\gamma]$ can be understood as follows. Note that a larger proportion of firms exerting effort in the fully optimal contract than the FPCR one for any fixed β in this segment. Further, as β goes up, the difference between proportions of firms exerting effort in two menus rises, though firms under both menus will exert more effort. Specifically, under the fully optimal and the optimal FPCR contracts, the proportions of firms who exert cost-reducing effort are $[\underline{\theta}, \underline{\theta}/(1-\beta)]$ and $[\underline{\theta}, 2\underline{\theta}/(2-\beta)]$, respectively, where the latter is a subset of the former. As β goes up, both sets are larger while $\underline{\theta}/(1-\beta)$ increases at a faster rate than $2\underline{\theta}/(2-\beta)$. The more effort-incentive essence of the fully optimal contract, therefore, could enlarge the welfare difference (a smaller G_F/G_O for a larger β) between these two menus. It is also worth noting that in the declining segment of G_F/G_O, the upper bound of the gain secured by the optimal FPCR is $3/4$ of the fully optimal one for any γ, which is the lower bound in [1].

When β is large, both the optimal FPCR and the fully optimal contract provide strong incentives such that *all* firms exert effort. More precisely, for any given θ the optimal cost-reducing effort induced by the optimal FPCR contract $e^*(\theta) = \beta\theta/(\beta+1)$ is closer to that of the fully optimal contract $e^*(\theta) = (\beta-1)\theta/(\beta+1) + \underline{\theta}/(\beta+1)$ as β increases. Thus the performance of the optimal FPCR contract approaches the fully optimal one and such a conclusion does not depend on the distribution of innate cost since for large β eventually all the firms exert some positive effort.

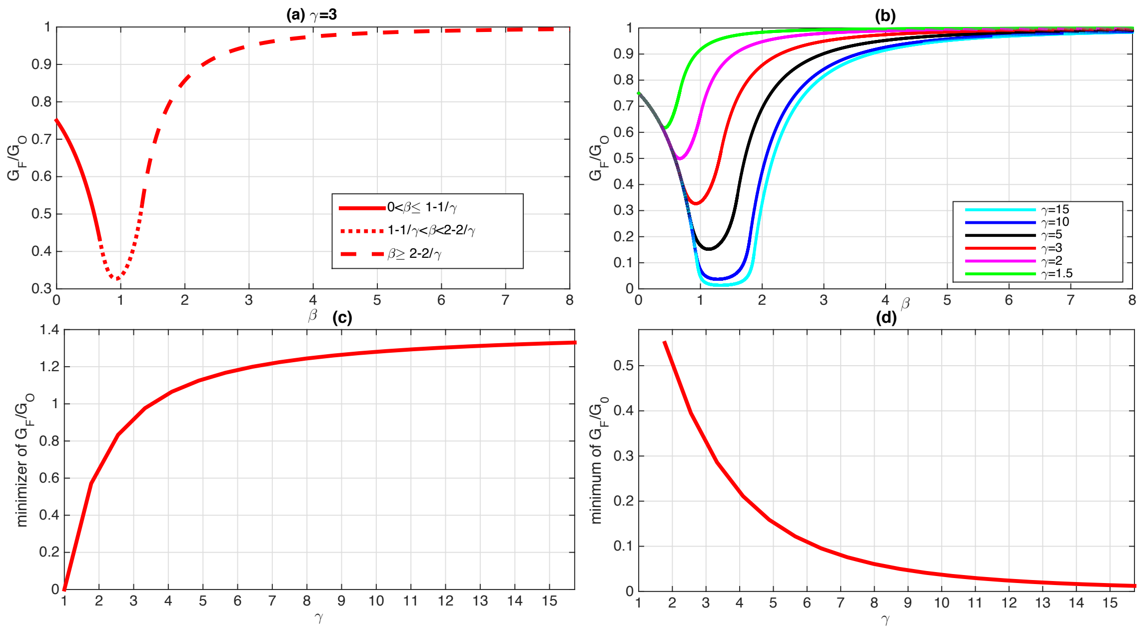

Figure 2.1: Illustration of performance

Theorem 2. *For any given* $\beta \in (1-1/\gamma, 2-2/\gamma)$*,* G_F/G_O *is strictly decreasing in* γ *and* $\lim_{\gamma\to\infty} G_F/G_O = 0$*.*

The result in Theorem 2 describes the effects of γ on the efficiency of the FPCR menu. The finding is of special interest because it implies that for any given $\beta \in (1-1/\gamma, 2-2/\gamma)$ and for any $\epsilon > 0$, there exists a $\gamma_0 > 1$ such that for any $\gamma \geq \gamma_0, G_F/G_O \leq \epsilon$, i.e., if the firms' innate costs are sufficiently spread, the efficiency of the FPCR menu would be arbitrarily close to that of a CR contract. Theorem 2 is illustrated in panel (b) of Figure 2.1, where the pattern of G_F/G_O is similar across different γ but the line is lower for a larger γ. This result can be interpreted using the conclusion in Lemma 1, which implies that the ratio of the proportion of firms choosing CR contracts to the one choosing FP contracts is

$(2/\beta - 1)\gamma - 2/\beta$. The ratio is strictly increasing in γ, and this indicates that if agents' innate costs are sufficiently dispersed, the firms choosing CR contracts would dominate and the performance of the FPCR contract would be close to a CR contract.

Our result is in contrast with that of [1], where the ratio G_F/G_O is bounded below at $3/4$ for any given γ.[2] The main force that leads to this discrepancy is that under a convex function, the optimal cost-reducing effort exerted by the agent is strictly increasing in her innate cost and a cost function with higher marginal cost induces larger cost-reducing effort for a given innate cost of the agent. If cost function is an identity, however, the optimal cost-reducing effort is a constant, regardless of the innate cost of the agent. The results in Theorems 1-2 are of particular importance for the principal to design a FPCR menu. It suggests that the FPCR menu is preferable when the marginal cost of agents is sufficiently large relative to the marginal disutility, whereas the menu is less appealing when the marginal cost is small and agents have more dispersed innate costs.

By further exploring the impacts of γ on G_F/G_O, we find that the minimizer and the minimum of G_F/G_O are strictly increasing and decreasing in γ, respectively. Unfortunately, a tractably analytical expression of the minimizer and minimum are not available, we instead employ numerical results to illustrate how they vary with γ, as shown in panels (c) and (d).

2.4 Conclusions

We extended the FPCR menu to allow firms' cost to be convex in their types and found that the performance of the optimal FPCR contract relies crucially on the cost function. On the one hand, the expected surplus can be very close to that of the fully optimal contract if the marginal cost is large. On the other hand, if the marginal cost is small and firms are diverse in their innate costs then the optimal FPCR menu behaves arbitrarily close to

[2] The parameter γ in [1] is defined as the ratio of $\bar{\theta} - \underline{\theta}$ to a constant.

a cost-reimbursement contract. Our findings suggest that the information on cost structure of firms is essential to the principal when the powerful FPCR is implemented.

3. ECONOMETRICS OF MULTI-PERIOD SIMPLE CONTRACTS

3.1 Introduction

Due to the fundamental role it plays in the studies of informational asymmetries and incentives, contract theory has attracted much attention from economists during the past three decades (see [15]; and [16]). One branch of the contracts are the *complex optimal contracts* in the spirit of [2] where the optimal payment to agents is a nonlinear function of both agents' unobserved type and their observed cost. Nevertheless, recent studies argue that another branch, the *simple menus of contracts* which oftentimes specify the payment only as a function of the agents' observed cost or even as a constant, could be more useful in practice (e.g., see [17]). Theoretical and empirical evidence show that these simple menus could capture a substantial proportion of the surplus that complex optimal nonlinear contracts would achieve ([1]).

Despite the importance of simple contracts and their wide use in various sectors, rigorous econometric analyses on this large class of contracts are largely missing in the literature. In actuality, rigorous identification results were developed only recently for complex optimal nonlinear contracts in [3, 5]. However, the methodology for complex contracts do not directly apply to simple contracts. This is because simple menus of contracts have different implications from complex contracts on how asymmetric information and incentives govern the relationship between a principal and an agent (e.g., see [1]; [14]). The fact that simple contracts are often implemented for multiple periods ([18]) imposes further challenges to rigorous econometric analysis for multi-period simple contracts because renegotiation and information revelation complicate the relationship between observables and model primitives.

This paper provides the first set of results on rigorous identification of multi-period simple contracts by focusing on a widely-used menu “fixed-price-cost-reimbursement (FPCR)" of simple contracts.[1] This menu consists of a fixed-price (FP) contract, in which the payment to the agent is a fixed price, regardless of the agent’s realized cost; and a cost-reimbursement (CR) contract, in which the agent is reimbursed exactly for all the realized cost. A multi-period FPCR contract may be implemented in two forms: a contract under renegotiation, which allows the principal and the agent to renegotiate on the initial contract at certain time during the implementation, or a contract under commitment, which prohibits any adjustment during implementation. A FPCR contract can be understood as a special case of a linear cost-sharing-cost-reimbursment (LCSCR) contract which has two options: a CR contract and a linear cost sharing (LCS) contract that specifies a lump-sum payment and a single fraction $\kappa \in [0, 1]$, of realized costs for which the agent would be reimbursed.

3.1.1 Model and Identification

We propose a general theoretical model for two-period FPCR contracts by extending the work of [18], which provides a simple framework for dynamic FPCR contracts. At the beginning of the first period, an agent chooses the most profitable contract from the FPCR menu provided by the principal for two periods. The agent is allowed to renegotiate with the principal and make changes to his initial choice at the end of the first period if the contract is renegotiable; whereas under commitment the agent has to stick to his initial choice. We derive the equilibrium conditions for the models under both commitment and renegotiation, where there are two and three segments of agents, respectively, and the agents within a segment make the same choice of contracts. In general, more efficient firms choose fixed-price contracts and exert cost-reducing effort while less efficient firms

[1]For example, many local authorities in France use this menu to contract with firms to provide the transport service. Other examples include the Indian customized software industry studied by [19], U.S. Air Force engine procurement considered in [17], and so on.

prefer cost-reimbursement contracts.

To conduct a rigorous econometric analysis of the model above, we provide constructive arguments to nonparametrically identify the structural elements in contracts under renegotiation. The argument is readily carried over to the commitment setting. Specifically, we show that the equilibrium conditions implied by the model, together with some exclusion restrictions to agents' types, enable us to nonparametrically identify the model primitives (if firms exert effort) from the observed information in a typical dynamic FPCR contract, i.e., the joint distribution of agents' choices of contracts, their realized costs, received payments from the principal. The model primitives include agents' cost and disutility functions, distribution of types, and two scalar parameters characterizing agents' bargaining power and the intertemporal preference.

Our identification argument takes several steps. First, by using the recently developed methodology in measurement errors ([20]) we recover the distribution of the unobserved optimal effort exerted by agents from the joint distribution of two covariates correlated with the effort. The one-to-one mapping between agent's observed cost and optimal effort implied by the equilibrium conditions further enables us to back out the (pseudo) optimal effort corresponding to each of the observed cost. Second, we rely on an exclusion restriction, i.e., the existence of some exclusive variables that directly affect the optimal effort but not the distribution of innate costs (types) to identify the cost structure of agents. The identification is achieved by exploiting variations of the quantiles for the cost when the exclusive variables change while the quantiles of the type remain the same. Third, we construct the pseudo innate costs from the identified cost function as well as the structural link between innate cost and optimal effort. Thus the distribution of innate cost is recovered. We then employ the structural elements identified above and the observed payment to the agent to recover the two parameters that characterize agents' bargaining power and intertemporal preference. Using the identification argument above, we also provide semiparametric identification results for the model if firms who do not exert effort are included in the model. Our identification results for multi-period FPCR contracts are without loss of generality and can be readily extended to LCSCR contracts or even some more general menus. The identification procedure also applies to FP and CR contracts since the FPCR contracts nests FP and CR ones as special cases. Based on the argument of identification, we also propose feasible procedures to estimate the model.

3.1.2 Preview of Empirical Findings

We apply our method to study the dynamic transport procurement contracts in France. The objective of the empirical study is to estimate the cost and disutility functions of agents, and other parameters of the local governments' (principal) and agents' preference. More importantly, we utilize these estimates to conduct a counterfactual analysis on the welfare comparison between FPCR contracts under renegotiation and under commitment.

Our empirical results reveal that both cost and disutility functions of agents are convex respectively in innate cost and effort, while identity cost function is generally assumed in the existing literature. We then employ the estimates to conduct an important counterfactual analysis for the purpose of comparing the welfare of FPCR contracts under commitment and under renegotiation. Our estimate shows that contracts under commitment achieve significant welfare gains comparing with those under renegotiation, this confirms the existing results in [18]. More importantly, we find new empirical evidence that both parties of the contract are welfare gainers: about sixty percent of the gains would accrue to taxpayers (the principal), whereas the firms (the agents) obtain the remaining forty percent. This is in contrast to [18], where only firms are welfare gainers.

Our new empirical evidence is mainly attributed to the convexity of firms' cost function since it implies that the marginal benefit of effort is increasing in innate cost, and in turn those inefficient firms (with higher innate costs) have stronger incentives to exert cost-reducing effort if they choose FP contracts. Besides, the empirical results show that a larger portion of firms choose FP contracts under commitment than under negotiation. Thus it is very likely that if we take both periods into account there would be more firms exerting cost-reducing effort under commitment and this creates higher welfare gains. In summary, our empirical findings indicate that it is crucial to take into account the functional form of firms' cost functions when one investigates the efficiency of contracts with incentives.

3.1.3 Relation to Existing Literature and Contributions

This paper contributes to a broader literature on the identification of contract models. There are few studies on the rigorous identification of contract models. Notable exception is [3], which show the nonparametric identification of a static *complex* contract model tailored from the seminal paper [2]. The identification argument in [3] does not apply to the simple contracts considered in this paper for the following reasons. First of all, the observed payment is (agents') type-specific in their complex contracts. Such payment mechanism provides richer variations for identification than the simple contracts, in which payments are oftentimes constant regardless of agents' type. Second, the model in our paper is multi-period. Renegotiation and information revelation in such contracts further complicate the behaviors of the principal and agents. It is not clear how one extends the identification results in [3] for a static model to its multi-period counterparts.

The identification in [3] relies on a one-to-one mapping between the observed price of the product and the private type of the agent, whereas such relationship does not exist in our model. Instead, we take advantage of newly developed results in measurement errors to recover the distribution of the unobserved optimal effort, and then identify the model structure by using an exclusion restriction. Identifying structural models through measurement errors has been widely in the literature, e.g., [21] and [22] provide general identification results for nonlinear models with misclassification and measurement error, respectively. Nevertheless, to the best of our knowledge, our paper is the first to employ the results in measurement errors to identify of contract models. In actuality, our identification argument developed for the two-period FPCR regulatory contracts could apply to a broader class of moral hazard and adverse selection models where agents with continuum of types are only offered a few simple contracts by the principal, e.g., labor contracts and insurance models.

In addition, our paper contributes to a growing empirical literature on contract theory (e.g.,[6] and [18]). [18] use the same source of data as ours to estimate the structural elements of dynamic FPCR contracts and compare the welfare of contracts under renegotiation and commitment. As described before, we estimate a richer model using different methods and find new empirical evidence on FPCR contracts. The significant convexity of the cost function implies that a linear cost specification in prior literature might be misleading. More importantly, we find that the convexity of firms' cost may lead to quantitatively different performance of FPCR contracts.

3.1.4 Roadmap

The rest of the paper is as follows. Section 2 presents the general dynamic FPCR model. In section 3 we establish the main identification results, and provide some discussions on the feasible estimation procedure. Section 4 analyzes the French transport procurement contracts. Section 5 concludes. Proofs, figures and tables are collected in the appendix.

3.2 The Model

A risk-neutral principal wishes to procure a project from a risk-neutral agent by offering a two-period menu consisting of two types of contracts in each period: a cost sharing (CS) contract in which the payment is dependent upon the agent's realized cost; and a cost-reimbursement (CR) contract in which the agent is reimbursed exactly for the realized cost. The essential difference between these two types of contracts lies in that CS contracts are high-powered incentive for cost reduction from efforts exerted by agents, while CR contracts are low-powered and provide no incentives for cost reduction.

The two-period setting introduces dynamics in terms of commitment and renegotiation. Obviously, the selection in the first contracting period partially reveals information on the firm's type: the choice of a CS contract signals that the firm is more efficient; the choice of a CR contract, however, conveys the information that the firm is less efficient. Under commitment, both parties are not allowed to renegotiate the initial contract or re-sign a new one even though new information is available. In the absence of commitment, however, both parties may be better off by exploiting these new information to renegotiate the initial contract. This lessens the parties' ex ante incentives when the prospect of renegotiation is perfectly anticipated in the contract design stage, thus eventually keeping parties from securing the efficiency that could have been obtained under commitment (e.g., [23]; and [24]).

At the beginning, nature randomly assigns the agent's innate cost (or "type") θ, which is distributed according to a cumulative distribution function $F(\cdot)$, with a density $f(\cdot)$ on a support $[\underline{\theta}, \bar{\theta}] \subset \mathbf{R}$. The agent observes its private value of θ, and the principal only has the knowledge of $F(\theta)$. The cost structure of realizing the project is

$$c_t \quad = \quad H(\theta - e_t), \qquad t \in \{1, 2\}$$

where c_t is the realized cost in period t, the innate cost θ represents the firm's management and production skills, which is invariant during two periods, and the private effort $e_t \geq 0$ which is unobserved by the principal, captures firm's actions taken to reduce cost c_t. The cost function $H(\cdot)$ takes a general form and nests the commonly assumed identity function as a special case.[2] The exerted effort e_t incurs some disutility according to a disutility function $\psi(e_t)$.

Given the principal's payment specification $p_t(c_t)$, the informational rent (profit) of the agent (firm) with type θ in period $t \in \{1, 2\}$ is defined as

$$U_t = p_t(c_t) - c_t - \psi(e_t) = p_t \circ H(\theta - e_t) - H(\theta - e_t) - \psi(e_t), \tag{3.1}$$

where $\circ$ denotes the composition of two functions. The firm's intertemporal profit is $U = rU_1 + (1-r)U_2$, where the weight r is defined as $r = 1/(1+\delta)$ with δ being the standard discount factor. The goal of the agent is to maximize U by choosing a contract from the CS-CR menu and then exerting the optimal effort e_t.

The principal designs the optimal menu of contracts by specifying the optimal form $p_t(c_t)$ for CS contracts in each period to maximize the expected social welfare $\mathbf{E}_\theta[SW(\theta)]$, where the expectation operator $\mathbf{E}_\theta$ is taken with respect to θ, and $SW(\theta)$ is the social welfare generated by the agent with innate cost θ:

$$SW(\theta) \equiv S - (1+\lambda)[rp_1 + (1-r)p_2] + \alpha[rU_1 + (1-r)U_2], \tag{3.2}$$

In the definition above, the dependence of the right side of (3.2) on θ is through the dependence of both U_1 and U_2 on θ. S is the gross surplus generated by the procured service and assumed to be sufficiently large to guarantee the desirability of the project. The cost of public funds $\lambda > 0$ captures some dead-weight loss due to a distortionary taxation for raising subsidies with the principal's intertemporal payment $rp_1 + (1-r)p_2$. The parameter α is the weight assigned to firms' profits by the principal ([25]; [26]), which can reflect the extent of the political pressure imposed by the agents on the political principals

[2] [4] also assume a general cost function in a monopoly model with regulation.

([27]) and therefore can be interpreted as firm's bargaining power against the principal in negotiation. We maintain that $\alpha < 1 + \lambda$, which captures the relevant trade-off between extracting profit and inducing efficient cost-reducing effort. Intuitively, the optimal menu of contracts offered by the principal trades off efficiency and rent extraction: CS contracts with large subsidy would induce the first-best effort while leaving much information rent to more efficient agents; CR contracts, however, nullify this rent without any incentive for agents to make cost-reducing effort.

In the following we assume that all related functions are at least twice continuously differentiable and that integration and differentiation can be interchanged. For a generic function $a(\cdot)$ with more than one argument, we denote its derivative with respect to the k-th argument by $a_k(\cdot)$.

Assumption 2. *(i) The random type θ is distributed according to $F(\cdot)$ with a density $f(\cdot) > 0$ on its support $[\underline{\theta}, \overline{\theta}] \subset \boldsymbol{R}$, $\underline{\theta} < \overline{\theta}$. (ii) $H(\cdot) \geq 0$, $H'(\cdot) > 0$, $H''(\cdot) > 0$. (iii) $\psi(\cdot) \geq 0$, $\psi'(\cdot) > 0$, $\psi''(\cdot) > 0$, $\psi(0) = 0$.*

Assumption 2 is standard in the procurement contract literature (e.g., [4]). Part (ii) suggests that an agent with a lower innate cost is more efficient, while an agent with a larger innate cost is less efficient. Part (iii) implies that the optimal effort under the CR contract is always zero. This is because under the CR contract, a firm is only reimbursed $c = H(\cdot)$ and $H(\theta) > H(\theta - e_t)$ for any $e_t > 0$, the optimal choice for the firm is to exert no effort. Therefore, its cost function reduces to $c = H(\theta)$ with the corresponding profit of firms being zero. Thus we focus mainly on the analysis of the CS contracts below whenever effort is involved.

Without loss of generality, in this paper we analyze a linear payment schedule, which is widely studied in the literature (e.g., [14]),

$$p_t = p_0 + \kappa c_t, \kappa \in [0, 1]. \tag{3.3}$$

Such a linear cost sharing (LCS) contract specifies a lump-sum payment p_0 and a single fraction $\kappa \in [0, 1]$ of realized cost for which the agent would be reimbursed. A special case of LCS contracts is the fixed-price (FP) contract in which the agent is paid a fixed-price p_0 whenever the principal specifies $\kappa = 0$ in equation (3.3). These FP contracts have not only attracted attention from theoerists (e.g., [1]; [14]; and [28]), they are also commmonly used in practice (e.g. [19]; and [17]). To accommodate our empirical application, we will discuss the theoretical properties and nonparametric identification for FPCR contracts (the principal provides a FP or CR contracts for an agent to choose from).

To begin with, we first consider the commitment setting as a baseline model. As is well known, commitment prevents both parties from renegotiating and thereby promotes efficient outcomes ex ante. As proved in [24], the equilibrium outcome under commitment is just the twice-repeated version of that in one-period (static) settings. Hence, we will focus on the analysis for the one-period contract and omit the period subscript t below.

3.2.1 Commitment

First, we consider the firm's problem. A firm with innate cost θ who chooses the FP contract maximizes its one-period profit by exerting the optimal effort e^*, which is determined by the first-order condition[3]

$$H'(\theta - e^*(\theta)) = \psi'(e^*(\theta)), \tag{3.4}$$

[3]Following [2] we assume that there exists a unique optimal effort for each innate cost.

that is, at the optimal level of effort the marginal benefit of exerting effort equals its marginal disutility. Moreover, the derivative of (3.4) with respect to θ on both sides leads to $\psi''(e^*(\theta))e^{*'}(\theta) = H''(\theta - e^*(\theta))(1 - e^{*'}(\theta))$, which implies that $0 < e^{*'}(\theta) < 1$ for all types θ who choose the FP contract under Assumption 2.

By incentive compatibility, if one type prefers a fixed-price contract, so do the more efficient types. This can be justified by the fact that $dU/d\theta = -H'(\theta - e^*(\theta)) < 0$. Accordingly, the support of firms' innate cost θ can be divided into two segments: more efficient agents choose the FP contract and less inefficient ones prefer the CR contract. Suppose the fixed price designed by the principal is b^C, the unique cut-off type θ^* is just indifferent between the two contracts such that

$$b^C = H(\theta^* - e^*(\theta^*)) + \psi(e^*(\theta^*)).$$

We next analyze the principal's problem. The goal of the principal is to maximize the expected social welfare by setting the unique best fixed price b^C

$$b^C \equiv \underset{b \in \mathbf{R}_+}{\text{argmax}} \ \mathbf{E}_\theta[SW(\theta, b)], \tag{3.5}$$

where

$$\begin{aligned} \mathbf{E}_\theta[SW(\theta, b)] &= S - (1+\lambda)\left(bF(\theta^*(b)) + \int_{\theta^*(b)}^{\bar{\theta}} H(\theta)dF(\theta)\right) \\ &\quad + \alpha \int_{\underline{\theta}}^{\theta^*(b)} [b - H(\theta - e^*(\theta)) - \psi(e^*(\theta))]dF(\theta), \end{aligned}$$

with the first-order condition

$$\left(1 - \frac{\alpha}{1+\lambda}\right)\frac{F(\theta^*)}{f(\theta^*)} = \frac{H(\theta^*) - b^C}{H'(\theta^* - e^*(\theta^*))}.$$

The following proposition summarizes the optimal outcome of a two-period contract model under commitment.

Proposition 1. *In the economic environment under commitment, under Assumption 2, the two-period optimal fixed-prices* (b_1^C, b_2^C) *satisfy* $b_1^C = b_2^C \equiv b^C$ *with the corresponding cut-off type* θ^* *such that*

$$b^C = H(\theta^* - e^*(\theta^*)) + \psi(e^*(\theta^*)), \tag{3.6}$$

$$\left(1 - \frac{\alpha}{1+\lambda}\right)\frac{F(\theta^*)}{f(\theta^*)} = \frac{H(\theta^*) - b^C}{H'(\theta^* - e^*(\theta^*))}. \tag{3.7}$$

In both periods more efficient firms ($\theta \leq \theta^*$) *choose the same FP contract, while less efficient firms* ($\theta > \theta^*$) *operate under CR contracts. The optimal effort* $e^*(\theta)$ *associated with the FP contract satisfies* $0 < e^{*'}(\theta) < 1$.

The property $0 < e^{*'}(\theta) < 1$ implies that less efficient firms will exert more effort. This is in contrast to the constant optimal effort e^* for any type θ in most of the existing studies on the FPCR menu. Intuitively, an identity cost function implies that the marginal benefit of effort is the same for firms with different innate costs. Therefore, a firm, no matter what its innate cost is, has no incentive to exert more effort than others. Nevertheless, when $H(\cdot)$ is non-identity, say convex, a firm with a higher innate cost enjoys a larger reduction of cost than a lower-innate cost firm by exerting the same effort. Indeed, our empirical evidence lends support to the convexity of cost function.

3.2.2 Renegotiation

Unlike in the contracts under commitment, the revelation principal fails under renegotiation due to the fact that after private information is revealed at the end of the first period, as in [29], the principal can offer a new contract by renegotiation which may ex-post benefit both parties. To analyze the equilibrium of dynamic contracts under renegotiation, we resort to the renegotiation-proof principal by following the related literature.[4]

As opposed to the two-period dynamic contracts under commitment where the optimal choices of contracts are either (b^C, b^C) (FP contracts with the same payment b^C in both periods) or $(H(\theta), H(\theta))$ (CR contracts with the same payment $H(\theta)$ in both periods), an optimal choice of contracts under the renegotiation case may involve an additional option, i.e., a CR contract in the first period followed by a FP contract in the second period. That is, the agent may renegotiate to change the choice of contract made at the beginning of the first period. Taking into account the possibility of renegotiation, the principal provides three options for the firms: A two-period FP contract $C_1^R \equiv (b_1^R, b_2^R)$, denoted by FF contract, a first-period CR contract followed by a second-period FP contract $C_2^R \equiv (H(\theta), b_3^R)$, denoted by CF contract, and a two-period CR contract $C_3^R \equiv (H(\theta), H(\theta))$, denoted by CC contract, where $b_j^R, j \in \{1, 2, 3\}$ is the optimal fixed price in the corresponding options, and $H(\theta)$ indicates the CR contract.[5] By using the renegotiation-proof principal, we obtain the following proposition that characterizes the equilibrium outcome of the renegotiation-proof menu of contracts $C^R \equiv (b_1^R, b_2^R, b_3^R)$.

Proposition 2. *In the economic environment under renegotiation, under Assumption 2, the optimal renegotiation-proof menu of two-period contracts $C^R = (b_1^R, b_2^R, b_3^R)$ satisfy $b_1^R = b_2^R \equiv \underline{b}^R$ and $b_3^R \equiv \bar{b}^R$ with two cut-off types (θ_1^*, θ_2^*) such that $\theta_1^* < \theta_2^*$ and*

$$\bar{b}^R = H(\theta_2^* - e^*(\theta_2^*)) + \psi(e(\theta_2^*)), \tag{3.8}$$

$$\underline{b}^R = r[H(\theta_1^* - e^*(\theta_1^*)) + \psi(e^*(\theta_1^*))] + (1 - r)\bar{b}^R, \tag{3.9}$$

[4]Renegotiation-proof principal: if a contract is a perfect Bayesian equilibrium in which renegotiation occurs in equilibrium, then there exists a renegotiation-proof contract that achieves the same outcome. In other words, any long-term agreement which is renegotiatable could be replaced by another long-term contract with a second-period continuation equal to the renegotiated offer.

[5]The choice consisting of a first-period FP fixed price contract followed by a second-period CR contract is never optimal for a firm. This is because the profit from a FP contract is strictly positive almost surely while the profit with CR contracts is always zero.

$$\left(1 - \frac{\alpha}{1+\lambda}\right) \frac{F(\theta_2^*) - F(\theta_1^*)}{f(\theta_2^*)} = \frac{H(\theta_2^*) - \bar{b}^R}{H'(\theta_2^* - e^*(\theta_2^*))}. \tag{3.10}$$

The most efficient types within the lower subinterval $[\underline{\theta}, \theta_1^*]$ *choose* C_1^R*; the intermediate efficient ones within the intermediate subinterval* $(\theta_1^*, \theta_2^*]$ *choose* C_2^R*; and the least efficient ones within the larger subinterval* $(\theta_2^*, \bar{\theta}]$ *choose* C_3^R*. In particular, the optimal effort* $e^*(\theta)$ *associated with the FP contract satisfies* $0 < e^{*\prime}(\theta) < 1$ *in each period.*

Most of the equilibrium outcomes under renegotiation are analogous to that under commitment, e.g., type-dependent effort. In particular, under renegotiation C_1^R is more powered-incentive than C_2^R due to the fact that C_2^R includes only one-period effort-incentive FP contract. Moreover, $\bar{b}^R$ has the identical functional form as b^C, and $\underline{b}^R$ is a weighted sum of a real fixed price $\bar{b}^R$ and a seeming "fixed price" $H(\theta_1^* - e^*(\theta_1^*)) + \psi(e^*(\theta_1^*))$ with weights $1 - r$ and r, respectively. Combining the facts that $H'(\theta - e^*(\theta))(1 - e^*(\theta)) + \psi'(e^*(\theta))e^{*\prime}(\theta) > 0$, $r > 0$, and $\theta_2^* > \theta_1^*$, we obtain $\bar{b}^R > \underline{b}^R$, which extends a similar relationship in [18] to our general model. The intuition behind $\underline{b}^R < \bar{b}^R$ can be ascribed to the fact that fixed prices must be raised sufficiently to induce those intermediate efficient firms with the initial choice of CR contracts to switch to FP contracts when the information on the agent's innate cost is revealed after the first period, whereas most efficient firms would choose FP contracts at the beginning even they are paid a relatively low fixed price $\underline{b}^R$ in both periods.

3.3 Nonparametric Identification

This section presents the identification of the model primitives under renegotiation, and the insight of identification naturally carries over to the commitment setting. Hereafter in the paper, we use capital and small letters to indicate random variables and their realizations, respectively except for θ to follow the convention in the literature.

The data report realized cost $C \in [\underline{c}, \bar{c}]$, two fixed prices $(\underline{B}^R, \overline{B}^R)$, and two binary choices D_1, D_2 indicating the choice of FF or CF contract, respectively. Consequently the binary variable $1 - D_1 - D_2$ indicates the choice to CC contracts. The fact that the realized costs are associated with different contracts allows us to divide the support $[\underline{c}, \bar{c}]$ into two subintervals $[\underline{c}, c_H]$ and $[c_L, \bar{c}]$ with $c_L \leq c_H$. The two subintervals cover the costs of firms who choose FP contracts either in FF contracts for two periods or in CF contracts for the second period, and who choose CR contracts either in CC contracts for two periods or in CF contracts for the first period, respectively. Specifically, the model in Section 2 implies[6]

$$\underline{c} = H(\underline{\theta} - e^*(\underline{\theta})), c_L = H(\theta_1^*), c_H = H(\theta_2^* - e^*(\theta_2^*)), \bar{c} = H(\bar{\theta}).$$

We also observe a vector of exogenous variables $Y \equiv (Y_1, Y_2, \cdots, Y_d) \in \mathbf{R}^d$ that characterize the principal, firms, and/or contract. For example, in our empirical application of transport procurement contracts, Y can be local government's political preference, firms' number of employee, and the size of the transport lines, etc. All the model primitives may depend on Y or its subvector. For a given $Y = y$, our model primitives can be summarized as the innate cost distribution $F(\cdot|y)$, the cost function $H(\cdot, y)$, the disutility function $\psi(\cdot, y)$, the weight a principal puts on firms' profit $\alpha(y)$, the principal's cost of public funds $\lambda(y)$, and the intertemporal weight of firms $r = 1/(1+\delta)$, which is independent of Y. Since our identification argument will be conditional on $Y = y$, we suppress y whenever there is no ambiguity and denote our model structure as $\mathbf{S} \equiv [F(\cdot), H(\cdot), \psi(\cdot), \alpha, r]$, where λ is not included for reasons that will become clear. We maintain that the observed data are generated from the model primitives $\mathbf{S}$ and the equilibrium conditions presented in the preceding section are satisfied .

We first note that the parameters α and λ cannot be separately identified. The intuition lies in that only the ratio $\alpha/(1+\lambda)$ matters directly for the optimal FPCR menu designed

[6]The condition $c_L \leq c_H$, or equivalently $\theta_1^* \leq \theta_2^* - e^*(\theta_2^*)$ is directly testable from observed costs in the data.

by the principal (r.f. (3.7) and (3.10)).[7] Next, we show that at best we can identify the model up to scale of the innate cost θ. Intuitively, the fact that both the cost function $H(\cdot)$ and innate costs θ are unknown suggests that a transformation of $H(\cdot)$ of θ may lead to the same observed data. The lemma as follows justifies this conjecture by showing that an alternative model structure $\widetilde{\mathbf{S}} \equiv [\widetilde{F}, \widetilde{H}, \widetilde{\psi}, \alpha, r]$ where $\widetilde{F}(\cdot) = F(\cdot/\xi_1)$, $\widetilde{H}(\cdot) = H(\cdot/\xi_1)$, and $\widetilde{\psi}(\cdot) = \psi(\cdot/\xi_1)$ for some positive scalar ξ_1 is observationally equivalent to the structure **S** in the sense that they both lead to the same joint distribution of $(C, \underline{B}^R, \overline{B}^R, D_1, D_2)$.

Lemma 3. *(Observational Equivalence) Suppose two structures* $\mathbf{S} \equiv [F, H, \psi, \alpha, r]$ *and* $\widetilde{\mathbf{S}} \equiv [\widetilde{F}, \widetilde{H}, \widetilde{\psi}, \alpha, r]$ *both satisfy Assumption 2, then* **S** *and* $\widetilde{\mathbf{S}}$ *are observationally equivalent.*

The proof of Lemma 3 is in the Appendix. Intuitively, the observational equivalence arises from the fact that both the cost function $H(\cdot)$ and it argument (type) θ are unobservables. A linear transformation of the original type θ into a new type $\widetilde{\theta} = \xi_1\theta$ for some nonzero parameter ξ_1 together with an appropriate transformation of $H(\cdot)$ leads to the same realized cost. Analogously, we may adjust other model primitives correspondingly to rationalize other observables $(\underline{B}^R, \overline{B}^R, D_1, D_2)$. This lemma suggests that we need to impose at least some normalization of the type in order to identify the model. In what follows, we normalize the lower bound of the type $\underline{\theta} = \theta_0$, which is formally stated in Assumption 4 below. From now on, we denote the support of θ by $[\theta_0, \bar{\theta}]$, which is divided into three subintervals $[\theta_0, \theta_1^*]$, $[\theta_1^*, \theta_2^*]$ and $[\theta_2^*, \bar{\theta}]$, where firms with type in the first two subintervals choose FF and CF contracts, respectively, whereas the ones in the last subinterval choose CC contracts.[8]

[7] The separate identification of α and λ maybe impossible even under stronger parametric restrictions, e.g., [18].

[8] Note that both the location and scale need to be normalized for the type θ in the analysis of a static monopoly contract model by [3], while in our model only a scale normalization is necessary. This discrepancy is due to the fact that our contract model is essentially different from theirs in the sense that the FPCR menu includes two types of contracts and the variation of firms' choices provides additional information for identification of their type θ.

3.3.1 Identification of Optimal Effort

The first challenge of identification concerns the unobserved optimal effort and the existing identifying strategies of contract theory, such as [3], do not apply to our model. Specifically, the identification of static monopoly contract model in [3] also involves the unobserved effort. Nevertheless, in addition to the observed cost more observables (i.e., output and price) are available for identification. By contrast, the directly related data in our model only consist of one observed continuous random cost variable and two binary choice variables. Acknowledging this difficulty, we adopt a newly developed method in measurement error, i.e., [20] to back out the distribution of the optimal effort nonparametrically.

Assumption 3. *There exists a subvector of Y, denoted by (Y_1, Y_2) without loss of generality, that is related to the optimal effort E^*:*

$$
\begin{aligned}
Y_1 &= E^* + V_1, \\
Y_2 &= m(E^*) + V_2,
\end{aligned}
$$

where $m(\cdot)$ is an unknown function and (E^*, V_1, V_2) are mutually independent with $\boldsymbol{E}V_1 = \boldsymbol{E}V_2 = 0$.

In the assumption above, the observable Y_1 can be understood as a normalization of the unobserved effort level and Y_2 can be chosen very flexibly since $m(\cdot)$ can be a very general function. This assumption is much less restrictive than the existence of double measurements of a latent variable required in the identification of many structural models, e.g., in [30], where $m(\cdot)$ has to be an identity function. The optimal effort or “hidden action" in contract theory is generally unverifiable, but firms’ effort-related performance is oftentimes measurable which can be further used as measurements for the optimal effort.

For example, [31] discusses the plausibility of employing cost-related variables to infer firms' effort. In general, the multi-dimensional measurement of firms' R&D would be a good example of Y_1 and Y_2.

Lemma 4. *(Schennach and Hu, 2013) Suppose Assumption 3 holds. Then under some regularity conditions, both $m(\cdot)$ and the distribution of E^*, $F_E(\cdot)$ are nonparametrically identified from the joint distribution of Y_1 and Y_2.*

Theorem 1 in [20] proves that $m(\cdot)$ and $F_E(\cdot)$ are both nonparametrically identified from the joint distribution of Y_1 and Y_2 if (1) $m(\cdot)$ is not a linear function, or (2) $m(\cdot)$ is linear but E^* is not normally distributed. In addition to Assumption 3, several additional regularity conditions are required for the identification. Nevertheless these conditions impose no further restrictions to our model. The main idea of the identifying strategy is to investigate the higher-order moments (characteristic functions) for the joint distribution of Y_1 and Y_2, which provides sufficient information to secure identification of the function $m(\cdot)$ and distribution $F_E(\cdot)$. We omit the proof of this lemma and refer the interested readers to [20] for details.

Recall that the optimal effort is increasing in firms' innate cost $0 < e^{*'}(\theta) < 1$ for any $\theta \in [\theta_0, \theta_2^*]$. The inverse function theorem implies that the innate cost is a strictly increasing function of the optimal effort with $\theta'(e^*) > 1$. Consequently, the argument $\theta - e^*(\theta)$ in $H(\cdot)$ is strictly increasing in the optimal effort e^* since $d(\theta - e^*)/d\theta = 1 - e^{*'}(\theta) > 0$. With a slight abuse of notation, let $L(e^*) \equiv \theta(e^*) - e^*$ and then the observed cost for firms who choose FF or CF contracts, $c = H(\theta(e^*) - e^*)$ can be rewritten as

$$c = H(L(e^*)) \equiv \widetilde{L}(e^*), \tag{3.11}$$

with $\widetilde{L}'(\cdot) = H'(\cdot)L'(\cdot) > 0$. A direct implication of this result is the existence of a one-to-one mapping between the cost and the optimal effort. Therefore, we obtain the following important structural link:

$$F_C(c) = \Pr(\widetilde{L}(e^*) \le c) = \Pr(e^* \le \widetilde{L}^{-1}(c)) = F_E(\widetilde{L}^{-1}(c)) = F_E(e^*), \tag{3.12}$$

where $F_C(\cdot)$ is the cumulative distribution function of cost C. This relationship together with the identified distribution function of effort $F_E(\cdot)$ enables us to obtain the pseudo optimal effort for any corresponding observed cost c,

$$e^* = \begin{cases} F_E^{-1}(F_C(c)), & \text{if } c \in [\underline{c}, c_H], \\ 0, & \text{if } c \in (c_L, \bar{c}], \end{cases} \tag{3.13}$$

where we utilize all the fixed-price contracts in the identification above, including fixed-price contracts for both periods in FF contracts and fixed-price contracts for the second period in CF contracts.

3.3.2 Identification of Cost Function

In this step of identification, we recover the cost function by using an exclusion restriction. First, we assume that the vector of characteristics $Y = (Z, W) \subset (\mathbf{R}^{d_1}, \mathbf{R}^{d_2}), d = d_1 + d_2$, where Z are variables that could be correlated with firms' type θ (of course affect firm's effort, too) and other variables W do not affect θ but enter the disutility function directly, thus influencing the effort and cost of firms who choose FP contracts. By construction, the vector of measurement for effort (Y_1, Y_2) could be in Z. To emphasize the difference between Z and W, we include W in those model primitives whenever necessary while still suppressing Z provided there is no ambiguity. Without loss of generality (wlog), we assume $d_2 = 1$, i.e., W is a scalar.

Assumption 4. *Assume that* (i) *There exists a variable* W *such that* $F(\theta|W) = F(\theta)$, *whereas W affects disutility function $\psi(\cdot, W)$. (ii) The cross derivative of ψ satisfies $\psi_{12}(e^*(\theta_0, W), W) = 0$ and $\psi_{12}(e^*(\theta, W), W) < 0$, for any $\theta \in (\theta_0, \theta_2^*]$.*

Part (i) requires the existence of an exclusive variable, that is, the distribution of the innate cost does not rely on the variable W (as will be shown W can be continuous or discrete given it takes at least two values), but the disutility function does, which captures the heterogeneity of the disutility across firms when cost-reducing effort is exerted though firms may be similar in their managerial ability (type). As a result, the optimal effort would rely on W and hence the firm's "productivity" $\theta - e$ varies with W. In our empirical application of French public transport contracts, a suitable choice of W can be a dummy indicating whether a firm is publicly or privately owned: the ex ante managerial ability of both types of firms does not depend on the ownership. However, the disutility of exerting cost-reducing effort might be higher in a publicly owned firm because working extra hours in such a firm may get paid higher due to regulations of labor union, etc. The first requirement of Part (ii) presents a normalization condition. It implies that the marginal disutility of effort at the lower bound of innate cost is constant across W. The second requirement of Part (ii) is standard in the theoretical literature on contracts, e.g., in [32] the marginal disutility of effort is assumed to be strictly decreasing with W.

The approach of exclusion restriction has been widely used in identifying structural models. For example, [33] use such an approach to nonparametrically identify first-price auctions with risk-averse bidders. In actuality, the existence of such an exclusive variable W may be empirically testable in our model using the result in the lemma below. Wlog, suppose our sample of contracts is classified into two subsamples based on two different realizations w_1 and w_2 of W with $w_2 > w_1$. Let $F_{C_j}(\cdot)$ be the distribution of the realized cost C in the subsample j corresponding to $w_j, j = 1, 2$. Then we have the following results regarding the relationship between $F_{C_1}(\cdot)$ and $F_{C_2}(\cdot)$.

Lemma 5. *Suppose Assumptions 2-4. Then the distribution* $F_{C_1}(\cdot)$ *strictly first-order stochastically dominates* $F_{C_2}(\cdot)$*. That is,* $F_{C_1}(c) < F_{C_2}(c)$ *for all* $c \in (\underline{c}, c_{H,1}]$*, where* $c_{H,1} \equiv H(\theta_2^* - e^*(\theta_2^*, w_1))$ *is the cost of the least efficient firm with* $W = w_1$ *who chooses FF or CF contracts. In addition,* $c_{H,1} > c_{H,2} \equiv H(\theta_2^* - e^*(\theta_2^*, w_2))$*,* $\underline{c}_1 \equiv H(\underline{\theta} - e^*(\underline{\theta}, w_1)) = H(\underline{\theta} - e^*(\underline{\theta}, w_2)) \equiv \underline{c}_2 = \underline{c}$*.*

The proof of this lemma is incorporated in the proof of Proposition 3 in the appendix. Lemma 5 provides a link between the observed costs of the two subsamples conditional on two different realizations of the exclusive variable W. The link is especially useful because the stochastic dominance of cost distributions indicates the existence of relationship between the two underlying values of W after controlling for other characteristics.

We now turn to the nonparametric identification of $H(\cdot)$. The main idea is to employ the exclusion restriction to exploit variations of the quantiles for the cost when W changes while the corresponding quantiles of the innate cost distribution are unchanged. Recall that the observed cost $C = H(\theta - e^*(\theta, W))$ for $\theta \in [\theta_0, \theta_2^*]$, thus we have

$$\theta = e^*(\theta, W) + H^{-1}(C). \tag{3.14}$$

Let us further suppress W whenever there is no ambiguity. The equation above implies that for any $\tau \in [0, 1]$, we have $\theta(\tau) = e^*(\theta(\tau), W) + H^{-1}(C(\tau))$, where $\theta(\tau)$ denotes the τ-th quantile of type distribution $F(\cdot)$, $C(\tau)$ denotes the τ-th quantile of cost distribution $F_C(\cdot)$. For $W = w_j, j = 1, 2$, let $e_j(\tau) \equiv e^*(\theta(\tau), w_j)$ and $C_j(\tau)$ be the τ-th quantile of $F_C(\cdot)$ corresponding to w_j. The exclusion restriction condition implies the key compatibility condition

$$\theta(\tau) = e_1(\tau) + H^{-1}(C_1(\tau)) = e_2(\tau) + H^{-1}(C_2(\tau)). \tag{3.15}$$

Rearranging terms leads to

$$H^{-1}(C_1(\tau)) = H^{-1}(C_2(\tau)) + e_2(\tau) - e_1(\tau) = H^{-1}(C_2(\tau)) + \Delta e(\tau),$$

where $\Delta e(\tau) \equiv e_2(\tau) - e_1(\tau)$. By iterating these $\Delta e(\cdot)$ for a monotone sequence of τ, we can identify $H(\cdot)$ on the support $[\theta_0 - \underline{e}, \theta_2^* - e^*(\theta_2^*, w_1)]$, where $\underline{e} \equiv e_1(\theta_0, w_1) = e_2(\theta_0, w_2)$. Specifically, $H(\theta_0 - \underline{e}) = \underline{c}$, and $H^{-1}(x) = \theta_0 - \underline{e} + \sum_{t=0}^{\infty} \Delta e(\tau_t)$ for any $x \in (\underline{c}, c_{H,1}]$, where the *unique* sequence $\{\tau_t\}_{t=0}^{\infty}$ is constructed as follows. Due to the the continuity and the strictly increase of cost function, there exists a unique $\tau_0 \in (0, 1]$ such that $C_1(\tau_0) = x$. Since $x = C_1(\tau_0) > C_2(\tau_0) > \underline{c}$, similarly there exists a unique $\tau_1 \in (0, \tau_0)$ such that $C_1(\tau_1) = C_2(\tau_0)$. Continuing such a construction gives rise to a unique sequence $\{\tau_t\}_{t=0}^{\infty}$ such that it is strictly decreasing with $0 < \tau_t \leq 1$ and satisfies the nonlinear recursive relation $C_1(\tau_{t+1}) = C_2(\tau_t)$ with the initial condition $C_1(\tau_0) = x$.

Proposition 3. *Suppose Assumptions 2-4 hold. Then the cost function $H(\cdot)$ is identified on* $[\theta_0 - \underline{e}, \theta_2^* - e^*(\theta_2^*, W_1)]$.

This proposition constitutes the first effort to provide a positive identification result of cost function $H(\cdot)$ in a dynamic cost-based contract model. By contrast, [5] show that the cost function $H(\cdot)$ in a static monopoly contract model can not be identified. Our positive result is due to (i) the innovative result on identification of effort's distribution $F_E(\cdot)$ and (ii) then existence of an exclusion restriction.

3.3.3 Identification of Type Distribution and Other Parameters

Having identified the cost function $H(\cdot)$ on $[\theta_0 - \underline{e}, \theta_2^* - e^*(\theta_2^*, W_1)]$ by using all the fixed-price contracts, we now present the identification of innate cost distribution $F(\cdot)$, the ratio $\alpha/(1+\lambda)$, the disutility function $\psi(\cdot)$ and firms' discount factor δ or equivalently their intertemporal weight r with $r = 1/(1+\delta)$.

The first step is to recover the pseudo-innate cost from its relationship with realized costs C. The basic idea is to construct a one-to-one mapping between θ and C by using $C = H(\theta - e^*(\theta))$, where $H(\cdot)$ is identified and e^* can be recovered for any given $c \in [\underline{c}, c_H]$ from (3.13). Recall that both FF and CF contracts are considered in the identification of $H(\cdot)$, the corresponding innate costs are on $[\theta_0, \theta_2^*]$. For all firms with innate costs in this interval, we combine (3.12) with (3.14) to get the pseudo-innate cost θ for the corresponding realized cost c,

$$\theta = H^{-1}(c) + e^* = H^{-1}(c) + F_E^{-1}(F_C(c)), c \in [\underline{c}, c_H], \tag{3.16}$$

where we dropped the variable W for ease of notation. Using the identified pseudo innate cost θ above, it is readily to recover a distribution $G(\cdot)$ and density function $g(\cdot)$ of θ on the interval $[\theta_0, \theta_2^*]$. Such an approach of identification has been widely used to identify structural models, e.g., in [34] the distribution of bidders' valuations are recovered from observed bids using a similar method. Nevertheless, it is important to notice that $g(\cdot)$ and $G(\cdot)$ are not the density function and CDF of firms' innate costs on $[\theta_0, \theta_2^*]$ yet, instead we have the following relationship between $g(\cdot)$ and $f(\cdot)$,

$$g(\theta) \equiv f(\theta | \theta_0 \leq \theta \leq \theta_2^*) = \frac{f(\theta)}{F(\theta_2^*)}, \forall \theta \in [\theta_0, \theta_2^*], \tag{3.17}$$

where $F(\theta_2^*) = \Pr(\theta_0 \leq \theta \leq \theta_2^*)$ is just the probability that a firm chooses FF or CF contracts. Therefore, $F(\theta_2^*) = \mathbf{E}(D_1 + D_2)$, where D_1 is defined as a binary variable indicating whether FF is chosen, and D_2 is defined similarly for CF contracts. The distribution of innate costs on the support $[\theta_0, \theta_2^*]$ can be recovered as

$$f(\theta) = g(\theta)\mathbf{E}(D_1 + D_2), F(\theta) = G(\theta)\mathbf{E}(D_1 + D_2); \forall \theta \in [\theta_0, \theta_2^*]. \tag{3.18}$$

Next we focus on identifying the ratio $\alpha/(1+\lambda)$ which describes the relative weight the principal puts on firms' profit and social cost. Since the principal's optimization problem involves both firms who choose fixed-price contracts and those choose cost-reimbursement ones, identification of the ratio requires information from all firms. We utilize the first-order-condition of the principal's problem (3.10) for identification. Note that $H'(\theta_2^* - e^*(\theta_2^*))$, $F(\theta_2^*)$ and $f(\theta_2^*)$ on the right-hand-side are identified,[9] it remains to to recover $F(\theta_1^*)$ and $H(\theta_2^*)$. First off, $F(\theta_1^*) = \Pr(\theta_0 \leq \theta \leq \theta_1^*)$ is just the probability that a firm chooses FF contracts, thus $F(\theta_1^*) = \mathbf{E}D_1$. Our analysis of the theoretical model in Section 2 shows that a firm with innate cost θ_2^* is indifferent between FP and CR contracts because both choices lead to zero profit.

$$\bar{b}^R - \Big(H(\theta_2^* - e^*(\theta_2^*)) + \psi(e(\theta_2^*))\Big) = mc - H(\theta_2^*) = 0, \tag{3.19}$$

where mc denotes the payment (realized cost) to the firm with innate cost θ_2^* and it is the lower bound of the realized costs for all firms who choose CC contracts since their innate costs are on $[\theta_2^*, \bar{\theta}]$ and $H'(\cdot) > 0$. Combining all the pieces above, we are able to identify the ratio $\alpha/(1+\lambda)$.

$$\frac{\alpha}{1+\lambda} = 1 - \frac{mc - \bar{b}^R}{H'(\theta_2^* - e^*(\theta_2^*))} \frac{f(\theta_2^*)}{\mathbf{E}D_2}. \tag{3.20}$$

Now we turn to the identification of the disutility function of effort $\psi(\cdot)$. Recall that we recovered the optimal effort as a function of realized costs in (3.13) for those costs in $[\underline{c}, c_H]$, i.e.,

$$e^* = F_E^{-1}(F_C(c)), \ \ c \in [\underline{c}, c_H].$$

[9] Due to the continuity of $H(\cdot)$, the derivative of $H'(\cdot)$ at the end point $\theta_2^* - e^*(\theta_2^*)$ can be identified as

$$H'(\theta_2^* - e^*(\theta_2^*)) = \lim_{\epsilon \to 0^-} \frac{H(\theta_2^* - e^*(\theta_2^*) + \epsilon) - H(\theta_2^* - e^*(\theta_2^*))}{\epsilon}.$$

By combining this relationship with (3.16), we obtain a one-to-one mapping between the optimal effort e^* and firms' type θ, $e^* = e^*(\theta)$ for all $\theta \in [\theta_0, \theta_2^*]$. This mapping enables us to identify the derivative of disutility function $\psi'(\cdot)$ from the first-order-condition of firms who exert optimal effort, i.e., $\psi'(e^*(\theta)) = H'(\theta - e^*(\theta))$ for all $\theta \in [\theta_0, \theta_2^*]$. An initial condition for this differential equation can be obtained using the first-order-condition (3.8), i.e.,

$$\psi(e)|_{e=e^*(\theta_2^*)} = \bar{b}^R - H(\theta_2^* - e^*(\theta_2^*)).$$

Thus the solution of $\psi(\cdot)$ is

$$\psi(e) = \bar{b}^R - H(\theta_2^* - e^*(\theta_2^*)) - \int_e^{e^*(\theta_2^*)} H'(e^{*-1}(v) - v)dv, \;\; e \in [\underline{e}, e^*(\theta_2^*)], \quad (3.21)$$

The last step is to identify firms' discount factor δ from (3.9),

$$\delta = \frac{\underline{b}^R - H(\theta_1^* - e^*(\theta_1^*)) - \psi(e^*(\theta_1^*))}{\bar{b}^R - \underline{b}^R}, \quad (3.22)$$

where θ_1^* is identified as $\theta_1^* = F^{-1}(\mathbf{E}D_1)$. The discount factor is a crucial objective to study agents' behavior and consequently conduct counterfactual or policy analyses. Nevertheless, a static or multi-period contract model with commitment does not allow us to recover δ. This can be clearly seen from (3.6) and (3.7) where agents' intertemporal behavior does not present. Actually, the discount factor oftentimes can not be identified in dynamic models. For example, [35] show that decision makers' discount factor in dynamic discrete choice models can not be identified.

Theorem 3. *Suppose Assumptions 2-4 hold. Then the principal's relative ratio $\alpha/(1+\lambda)$ and firms' discount factor δ are identified. The distribution of firms' innate cost $F(\cdot)$, disutility function $\psi(\cdot)$ and cost function $H(\cdot)$ are nonparametrically identified on $[\theta_0, \theta_2^*]$, $[e^*(\theta_0), e^*(\theta_2^*)]$ and $[\theta_0 - e^*(\theta_0), \theta_2^* - e^*(\theta_2^*)]$, respectively.*

Theorem 3 shows that the two-period FPCR model is nonparametrically identified for firms who exert effort at least in one period. The results can be readily applied to LCSCR contracts where the fixed-price contract is replaced by the linear cost-sharing contract. This is intuitive because LCS contracts provide additional variation of payments ($p_t = p_0 + \kappa c_t$) by comparing with FP contracts where the payment is a constant and independent of cost. In practice of procurement, a principal may only offer fixed-price contracts. In this case, our identification result applies directly, too. This is because our identification mainly relies on firms' incentives (effort), and excluding cost-reimbursement contracts does not affect our main argument of identification.

Now we consider the identification of CC contracts. Notice that the distribution of innate costs and cost function for firms who choose CC contracts are not identified in Theorem 3. This is because for firms who choose CR contracts in both periods, the only relation implied by the model is $c = H(\theta)$ where c is observed but both $H(\cdot)$ and θ are unknown. Furthermore, by nature of the FPCR contract, the firms' type is independent and so do their choices. Thus the model implications for FP contracts do not help us identify type distribution and cost function for CR contracts.

We show identification of the cost function $H(\cdot)$ on $(\theta_2^* - e^*(\theta_2^*), \bar{\theta}]$ and distribution of innate cost $F(\cdot)$ on $(\theta_2^*, \bar{\theta}]$ through imposing additional restrictions to the model. One of the choices for the restrictions is to parametrize the cost function $H(\cdot)$ while keeping $F(\cdot)$ and $\psi(\cdot)$ nonparametric. Specifically, we assume $H(\cdot)$ is known up to a finite dimensional parameter $\varrho_0 \in \mathbf{R}^L, L \geq 1$, denoted by $H(\cdot; \varrho_0)$ on its support. Under this parametric assumption, we modify the identification argument of Proposition 3 slightly to recover ϱ_0. We pick a sequence of quantiles $\tau_k \in (0, 1], k = 1, 2, \cdots, K, K \geq L$ to construct a nonlinear system

$$H^{-1}(C_1(\tau_k); \varrho_0) - H^{-1}(C_2(\tau_k); \varrho_0) - \Delta e(\tau_k) = 0, k = 1, 2, \cdots, K, \tag{3.23}$$

with further restrictions on the parameter $H'(\cdot;\varrho_0) > 0, H''(\cdot;\varrho_0) > 0$. Theorem 7 in [**?**] provides sufficient conditions under which global identification of ϱ_0 can be acheived. We skip the details and summarize the identification results in the following corollary by assuming ϱ_0 is globally identified.

Corollary 1. *Suppose Assumptions 1-3 hold and ϱ_0 is globally identified. Then the principal's relative ratio $\alpha/(1+\lambda)$ and firms' discount factor δ are identified. The distribution of firms' innate cost $F(\cdot)$ and disutility function $\psi(\cdot)$ are nonparametrically identified on $[\theta_0, \bar{\theta}]$ and $[e^*(\theta_0), e^*(\theta_2^*)]$, respectively.*

An alternative method to identify the full model is to parameterize only the distribution of firms' innate costs as $F(\cdot;\zeta_0)$. In this case, we follow the identification argument in Theorem 3 and the parameter ζ_0 would be identified after we recover the pseudo innate costs θ using (3.16). To identify the cost function $H(\cdot)$ on $(\theta_2^* - e^*(\theta_2^*), \bar{\theta}]$, we consider the realized costs for firms who choose CR contracts at least for one-period, i.e., $c_L < c < \bar{c}$,

$$\begin{aligned} F_C(c|c_L < c < \bar{c}) &= \Pr(H(\theta) < c|c_L < c < \bar{c}) = \Pr(\theta < H^{-1}(c) \mid \theta_1^* < \theta < \bar{\theta}) \\ &= F(H^{-1}(c);\zeta_0 \mid \theta_1^* < \theta < \bar{\theta}) \\ &= \frac{F(H^{-1}(c);\zeta_0)}{1 - F(\theta_1^*, \zeta_0)}. \end{aligned} \tag{3.24}$$

Considering that $H(\cdot)$ is strictly increasing, the equation above allows us to identify $H(\theta)$ on $(\theta_1^*, \bar{\theta}]$.

Note that the identification arguments in both Theorem 1 and Corollary 1 can be readily applied to FPCR contracts under commitment. Nevertheless, as we discussed earlier the discount factor r can not be identified.

3.3.4 Discussion on Estimation

The identification procedure of Theorem 1 is constructive and one may estimate the model following the identification steps. Nevertheless, a fully nonparametric estimating strategy is known to be data-demanding and thus could be of less interest in empirical studies.[10] Instead, we exploit the identification results of Corollary 1 to propose a semi-parametric procedure to estimate the model primitives in several steps, and leave the comprehensive analysis of asymptotic properties of estimators for future research.

Suppose our sample contains n contracts and n_f of them are FF and CF, the remaining n_c contracts are CC. We observe firms' realized costs $c_i, i = 1, 2, \cdots, n$, two measurements of effort $y_{1,i}$ and $y_{2,i}, i = 1, 2, \cdots, n_f$, two fixed-prices $\bar{b}^R$ and $\underline{b}^R$, and the exclusion variable W taking values of w_1 and w_2. Among the n_f contracts, n_1 and n_2 of them have $W = w_1$ and $W = w_2$, respectively.

First, we adopt a sieve maximum likelihood estimator proposed in [**?**] to estimate the density function of the optimal effort:

$$(\hat{f}_E, \hat{f}_{v_1}, \hat{f}_{v_2}, \hat{m}) = \text{argmax} \sup_{(f_E, f_{v_1}, f_{v_2}, m)} \frac{1}{n_f} \sum_{i=1}^{n_f} \ln \int f_{v_1}(y_{1,i} - v) f_{v_2}(y_{2,i} - m(v)) f_E(v) dv, \tag{3.25}$$

where the max and sup are taken over suitably restricted sets of functions; $f_{v_1}(\cdot)$, $f_{v_2}(\cdot)$, and $f_E(\cdot)$, respectively, denote the densities of error terms v_1, v_2 and optimal effort e^*. The optimization is subject to some restrictions which consist of constraints that the densities integrate to one and zero-mean constraints on the error densities $f_{v_1}(\cdot)$ and $f_{v_2}(\cdot)$. All four unknown functions $m(\cdot)$, $f_{v_1}(\cdot)$, $f_{v_2}(\cdot)$, and $f_E(\cdot)$ are chosen in an appropriate sieve space constructed by truncated series such as Hermite orthogonal series with the number of terms in the series increasing with the sample size.

[10] The essential step of the nonparametric estimation is to estimate the cost function $H(\cdot)$ nonparametrically. This can be done by following the procedure proposed in [33].

With the estimate of $f_E(e^*)$, one can obtain the estimate of $F_E(e^*)$ by integral

$$\widehat{F}_E(e^*) = \int_{\underline{e}}^{e^*} \widehat{f}_E(u)du.$$

Then we estimate the corresponding optimal effort for all $c \in [\underline{c}, c_H]$ by

$$\widehat{e}_i^* = \underset{e \in \mathbf{R}_+}{\operatorname{argmin}} \left| \widehat{F}_E(e) - \widehat{F}_C(c_i) \right|, \tag{3.26}$$

where $\widehat{F}_C(\cdot)$ is the empirical CDF of cost C. Applying the procedure above conditional on $W = w_1$ and $W = w_2$, we obtain $\widehat{F}_E(\cdot|w_j), j = 1, 2$ and shorten them as $\widehat{F}_{E,1}$ and $\widehat{F}_{E,2}$. The corresponding pseudo efforts are denoted $\widehat{e}_{i,1}^*, i = 1, 2, \cdots, n_1$ and $\widehat{e}_{i,2}^*, i = 1, 2, \cdots, n_2$. Analogously, we obtain $\widehat{F}_{C_1}(\cdot)$ and $\widehat{F}_{C_2}(\cdot)$ and denote the realized costs in the two subsamples as $c_{i,1}$ and $c_{i,2}$.

Next, we discuss how to estimate the parameter ϱ_0 following the identification argument in Section 3.2. Deviating slightly from the identification, we parametrize $H^{-1}(\cdot)$ as $H^{-1}(\cdot; \varrho_0)$ for the purpose of exposition. For the first subsample with $W = w_1$, we construct n_1 quantiles $\{\tau_{i,1}\}, i = 1, 2, \cdots, n_1$ using the realized costs $c_{i,1}$: $\tau_{i,1} = \widehat{F}_{C_1}(c_{i,1})$. The corresponding sequence of effort $\widehat{e}_{i,1}^*$ is obtained from $c_{i,1}$ by using (3.26) for $i = 1, 2, \cdots, n_1$. By plugging these quantiles to $\widehat{F}_{E,2}$, we obtain a sequence of pseudo effort $\{\tilde{e}_{i,2}^*\}$ such that $\tau_{i,1} = \widehat{F}_{E,2}(\tilde{e}_{i,2}^*), , i = 1, 2, \cdots, n_1$. Next, we proceed to construct n_1 corresponding costs $\tilde{c}_{i,2}$ for the second subsample such that $\tau_{i,1} = \widehat{F}_{C_2}(\tilde{c}_{i,2})$. We use $\tilde{e}_{i,2}^*$ and $\tilde{c}_{i,2}$ to emphasize that they are not necessarily in the sample. The identification argument provides the following nonlinear equations

$$H^{-1}(c_{i,1}; \varrho_0) - H^{-1}(\tilde{c}_{i,2}; \varrho_0) = \tilde{e}_{i,2}^* - \widehat{e}_{i,1}^*. \tag{3.27}$$

Analogously, we start from the second subsample to construct $\{\tau_{i,2}\}, \{\widehat{e}_{i,2}^*\}$, $\{\tilde{e}_{i,1}^*\}$ and

$\{\tilde{c}_{i,1}\}, i = 1, 2, \cdots, n_2$. Consequently we have n_2 more equations similar to (3.27). The parameter ϱ_0 is estimated using a nonlinear least square estimator:

$$\widehat{\varrho}_0 = \underset{\varrho_0 \in \mathbf{R}^L}{\text{argmin}} \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} \Big(H^{-1}(c_{i,1}; \varrho_0) - H^{-1}(\tilde{c}_{i,2}; \varrho_0) - \tilde{e}^*_{i,2} + \widehat{e}^*_{i,1} \Big)^2 \right.$$
$$\left. + \frac{1}{n_2} \sum_{i=1}^{n_2} \Big(H^{-1}(\tilde{c}_{i,1}; \varrho_0) - H^{-1}(c_{i,2}; \varrho_0) - \widehat{e}^*_{i,2} + \tilde{e}^*_{i,1} \Big)^2 \right\}. \quad (3.28)$$

It follows that $\widehat{\theta}_i = H^{-1}(c_i; \widehat{\varrho}_0) + \widehat{e}^*_i$ for FF and CF contracts, and that $\widehat{\theta}_i = H^{-1}(c_i; \widehat{\varrho}_0)$ for CC contracts. Using the estimates $\{\widehat{\theta}_i\}_{i=1}^n$, we can get a kernel estimator of $f(\cdot)$ and an empirical estimator of $F(\cdot)$, denoted by $\widehat{f}(\cdot)$ and $\widehat{F}(\cdot)$, respectively.

To estimate the parameters $\alpha/(1+\lambda)$ and r, it is necessary to estimate the boundary points θ^*_1 and θ^*_2. We exemplify the estimation using θ^*_2 where all the firms with innate costs $\theta < \theta^*_2$ choose FP contract at least in one period. Thus the probability $\Pr(\theta < \theta^*_2) = F(\theta^*_2)$ can be approximated by the ratio n_f/n and this implies the estimator $\widehat{\theta}^*_2 = \arg\min_\theta \left|\widehat{F}(\theta) - n_f/n\right|$. The point $\widehat{\theta}^*_1$ can be analogously estimated. Using a minimum distance estimator, we obtain the estimates for $\alpha/(1+\lambda)$ and the weight r by plugging the estimates above into (3.10) and (3.9), respectively.

Lastly, we can estimate $\psi(\cdot)$ by the first-order condition $H'(\theta - e^*; \varrho_0) = \psi'(e^*)$. Due to the nonparametric relationship between the optimal effort and type, we need to approximate the integral of $H'(\cdot; \widehat{\varrho}_0)$ over a continuous interval by summation over a finite sequence of grids for the interval. To be specific, for any $e \in [\underline{e}, \widehat{\overline{e}}]$, where $\widehat{\overline{e}} \equiv \widehat{e}^*(\hat{\theta}^*_2)$, we split the interval $[e, \widehat{\overline{e}}]$ into $k > 1$ (k can vary with e) subintervals evenly $[e, e_1], [e_1, e_2], \cdots, [e_{k-1}, e_k]$ with $e_k = \widehat{\overline{e}}$. For each $j = 1, \cdots, k$, one obtains the corresponding estimate of c_j by $\widehat{c}_j = \arg\min_{c \in \mathbf{R}_+} |\widehat{F}_E(e_j) - \widehat{F}_C(c)|$ as well as the corresponding estimate of θ_j by $\widehat{\theta}_j = H^{-1}(\widehat{c}_j; \widehat{\varrho}_0) + e_j$. Based on the identification result (3.21), we can estimate $\psi(e)$ as

$$\widehat{\psi}(e) = \bar{b}^R - \widehat{c}_m - \frac{\widehat{\bar{e}} - e}{k} \sum_{j=1}^{k} H'(\widehat{\theta}_j - e_j; \widehat{\varrho}_0), \tag{3.29}$$

where $\widehat{c}_m = \max_{i=1,2,\cdots,n_f}\{c_i\}$ is an estimator of $H(\theta_2^* - e^*(\theta_2^*))$ and $H'(H^{-1}(\cdot), \hat{\varrho}_0) = 1/H^{-1\prime}(\cdot; \hat{\varrho}_0)$.

3.4 Empirical Application

In this section we apply our method to analyze transport procurement contracts in France. As in the theoretical model, the local authority (the principal) provides FP and CR contracts to procure public transport service from firms (agents). Regulatory rules require that these contracts must be renegotiated every five years between the two parties. Thus the dataset is particularly suitable for our model. [18] use the same source of data to analyze the dynamic FPCR contracts and conduct a welfare comparison between contracts under commitment and renegotiation. In addition to different estimation strategies, our application is also different from [18] in that we allow both agents' cost function and disutility function to be nonlinear while their cost function is simplified as an identity function and consequently disutility is constant across innate costs. Our empirical findings show that the nonlinearity of agents' cost function is crucial to the welfare analysis of FPCR contracts.

3.4.1 Data

The dataset includes 543 two-period contracts implemented from 1987 to 2001. Among these contracts, 281 observations are two-period fixed-price contracts (FF), 88 observations are CR contract in the first period followed by one FP contract in the second period (CF), and the remaining 174 ones are a two-period CR contracts (CC). For each contract, the dataset reports the type of contract, the realized cost, the subsidy (payment from the principal, i.e., the fixed prices in the FP regime), the network size (defined as the length of transport network specified in a contract), the labor fee, the political preference of local governments (a dummy variable with 1 indicating right-wing and 0 left-wing), and some characteristics of firms: the number of employees, the number of drivers, the size rolling stock (measured by the number of vehicles), and the groups that own the firms.

Table 3.1 provides a summary statistics of our data. On average the cost is about 17 million euros and the subsidy is approximately 19 million per contract, thus indicating on average the firms are profitable. The average labor fee is 10.7 million and accounts for 64 percent of the total cost, suggesting that reducing the labor fee is critical to increase the firm's profit. The average number of employees is 413, thereby implying the intensive labor of transport industry. The network size ranges from 58km to 1184km with average 203km and median 236km, thus showing the symmetry of the distribution of this covariate. The major four groups owning most of transport firms are: Connex (25 percent), Agir (20 percent), Transdev (19 percent), and Keolis (30 percent) with their respective market shares in parenthesis. $Others$ in Table 3.1 are the small groups accounting for the rest 5 percent of firms which are not included in our data. A local government is classified as right- and left- wing according to its political preference. From the data, 52 percent of the local governments are right-wing, implying that the distribution of political preference is highly symmetric, which may result from some political equilibrium.

3.4.2 Empirical Strategy

The two primary goals of our application are (1) to estimate the cost and disutility functions and test the commonly imposed restrictions on them in the existing literature; (2) to assess the welfare gains that would be achieved if both parties commit to long-run contracts, as well as the distribution of those gains between the principal and firms. The estimation takes three steps: First, we estimate the parameters in the cost and disutility functions. Second, we recover the two cut-off types which are modeled to be dependent on

Table 3.1: Summary statistics

Variables	Mean	Std. Dev.	Min	Median	Max
# of Contracts	543				
# of FF	281				
# of CF	88				
# of CC	174				
Cost	16860	15954	2397	10347	93993
Subsidy	18794	18236	2265	12039	114483
Number of employees	413	364	68	267	1772
Labor fee	10740	10241	716	6650	53178
Rolling stock	165.48	121.41	33	83.5	724
Drivers	278.16	215.51	47	144.2	1181.5
Network size	294	203	58	236	1184
Connex	0.20				
Agir	0.20				
Tran	0.27				
Keolis	0.28				
Others	0.05				
Right wing	0.52	0.50	0.00	1.00	1.00

* All variables are real terms. The units of cost, subsidy, and labor fee are 1000 euros.

firms' characteristics. Finally, we estimate the distribution of the innate cost and bargaining powers. Recall that the identification argument in Section 3 is presented conditional on given characteristics. In comparison, we consider in the current section a general environment that allows rich heterogeneity among the principal and firms. Acknowledging the limited sample size, we adopt a parametric specification of model primitives to make estimation feasible.

3.4.2.1 Cost and Disutility Functions

Recall that we presented in Section 3 a nonparametric approach to recover the pseudo innate cost by first estimating the distribution of effort through the joint distribution of two related variables. A practical issue of such a nonparametric approach in our application is that a large sample size is required due to the slow rate of convergence for the sieve estimator $\widehat{f}_E(\cdot)$ in (3.25).[11] Thus we take the alternative approach by parameterizing the

[11] Please see [22] for the detailed asymptotic properties for the estimator.

model primitives.

We specify the cost function under the FP scheme by taking into account firms' heterogeneities

$$c \;=\; H^F(\theta - e^*, z) = \beta_1(\theta - e^*) + \beta_2(\theta - e^*)^2 + z'\beta_3. \tag{3.30}$$

where z is a vector capturing the characteristics of the principal, the firms as well as the contract, and adopt the following homogenous disutility function $\psi(\cdot)$.

$$\psi(e) \;=\; \gamma_1 e + \gamma_2 e^2, \tag{3.31}$$

which satisfies the normalization condition $\psi(0) = 0$. This normalization has been adopted by related empirical studies on cost-based contracts, such as [6]. As noted, the covariate W is not included in the disutility function as required in our identification. This is mainly because our parametric specifications greatly relax the restrictions imposed for nonparametric identification and the dependence on W is not necessary any more. According to the specifications above and the equilibrium condition in Section 3.2, the optimal effort is

$$e^*(\theta) = \frac{\beta_1 - \gamma_1}{2(\beta_2 + \gamma_2)} + \frac{\beta_2}{\beta_2 + \gamma_2}\theta, \tag{3.32}$$

with $e^{*'}(\theta) \in (0, 1)$.

Under the parametric specification, we need only one measurement for the optimal effort to identify the model parameters instead of two measurements for optimal effort in the nonparametric identification. Suppose there is a measurement x for the unobserved optimal effort through $x = d_0 + d_1 e^* + \eta$, where η is an i.i.d. error with conditional zero mean $E[\eta|x, z] = 0$ and unknown constant variance σ_η^2 (to simplify exposition, we drop the index of observation i in the notation). The functional form of $e^*(\theta)$ in (3.32) implies that following relationship between firm's innate cost θ and the measurement x:

$$\theta = \widetilde{\rho}_0 + \widetilde{\rho}_1 x + \widetilde{\eta} \tag{3.33}$$

where $\widetilde{\rho}_0 = [d_1(\gamma_1 - \beta_1) - 2d_0(\beta_2 + \gamma_2)]/(2d_1\beta_2)$, $\widetilde{\rho}_1 = (\beta_2 + \gamma_2)/(d_1\beta_2)$, and $\widetilde{\eta} = -(\beta_2 + \gamma_2)\eta/(d_1\beta_2)$. The relationship between x and θ in (3.33) can apply to both FP and CR contracts based on the one-to-one mapping between the optimal effort and the innate cost. The approach of measurement has strong empirical support, at least for the transport industry in France, where the number of employee is found to be one of the most significant input to measure firms' efficiency (e.g., [36]. In our data, the average labor fee is about 64 percent of the total cost, which also indicates the importance of number of employees to measure firms' efficiency. Following these results, we choose the number of employees as the measurement for optimal effort (and hence innate cost).

Analogously, the cost function under the CR scheme is specified as

$$c \;\; = \;\; H^C(\theta, z) = \beta_1\theta + \beta_2\theta^2 + z'\beta_4, \tag{3.34}$$

where we allow for possibly different marginal costs of characteristics by introducing two parameter vectors β_3 and β_4 in (3.30) and (3.34), respectively. By plugging (3.33) into 3.34 it is easy to show that $(\widetilde{\rho}_0, \widetilde{\rho}_1, \beta_1, \beta_2, \beta_4)$ can not be identified. Then we can normalize $\widetilde{\rho}_0 = 0$ and $\widetilde{\rho}_1 = 1$ based on the intuition that this can be interpreted as normalization of innate costs.[12] Indeed, the normalization of innate costs is also necessary in the nonparametric

[12]This normalization implies that $d_0 = (\gamma_1 - \beta_1)/(2\beta_2)$ and $d_1 = (\beta_2 + \gamma_2)/\beta_2$.

identification. Then by substituting x for θ in (3.34), we obtain[13]

$$c = \widetilde{\beta}_0 + \beta_1 x + \beta_2 x^2 + z'\beta_4 + \varepsilon_1, \tag{3.35}$$

where $\widetilde{\beta}_0 = \beta_2\sigma_\eta^2$, and the composite error $\varepsilon_1 \equiv \beta_1\eta + 2\beta_2 x\eta$ satisfies $E[\varepsilon_1|x,z] = 0$ under $E[\eta|x,z] = 0$. Similarly, by substituting this measurement for optimal effort into (3.30) gives rise to the realized cost for FP contracts:

$$c = \widetilde{\gamma}_0 + \widetilde{\gamma}_1 x + \widetilde{\gamma}_2 x^2 + z'\beta_3 + \varepsilon_2, \tag{3.36}$$

where the composite parameters $(\widetilde{\gamma}_0, \widetilde{\gamma}_1, \widetilde{\gamma}_2)$ are explicitly parametric functions of the primitive parameters $(\beta_1, \beta_2, \gamma_1, \gamma_2)$ and the details including the composite error ε_2 are provided in the appendix. Under the assumption $E[\eta|x,z] = 0$, it follows $E[\varepsilon_2|x,z] = 0$.

The parameters $(\beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2)$ are estimated as follows. We first use the CC contracts to estimate the parameters in (3.35), and then use CF (the fixed-price contracts, i.e., the second period) and FF contracts to recover parameters in (3.36). By combing these estimates, we obtain $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\gamma}_1, \hat{\gamma}_2)$.

3.4.2.2 Distribution of Innate Costs

Employing the equilibrium conditions (3.8) and (3.9), we are able to express the observed fixed prices $\bar{b}^R$ and $\underline{b}^R$ in CF and FF contracts, respectively, as follows.

$$\begin{aligned} \bar{b}^R &= \rho_0(\gamma_1, \gamma_2, \beta_1, \beta_2; \theta_2^*, z) + z'\beta_3 + \varepsilon_3, \\ \underline{b}^R &= \rho_1(\gamma_1, \gamma_2, \beta_1, \beta_2, r; \theta_1^*, z) + z'\beta_3 + \varepsilon_4, \end{aligned} \tag{3.37}$$

[13] Our cost specification here does not involve the interaction terms $z'\theta$, since none of these interaction terms when added into the cost specification is significant at the 10 percent level. Hence, the marginal cost of each element in z does not rely on its type θ.

where θ_1^* and θ_2^* are the cut-off types in proposition 2. The expression of $\rho_0(\cdot)$ and $\rho_1(\cdot)$ are provided in the appendix. We introduce some structural errors ε_3 and ε_4 to account for other unobserved factors that may affect the fixed prices, and parametrize $\theta_k^*, k = 1, 2$ as $\theta_k^* = z^{'}\pi_k$ to model their dependence on characteristics. Then we use observations for CF and FF contracts to estimate the parameters π_1, π_2, and r by a linear regression, where the estimates $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\gamma}_1, \hat{\gamma}_2)$ are from the previous step.

Next we specify the distribution of firms' type θ as a normal with mean $\mu(z)$ and variance σ^2, where $\mu(z) \equiv z'\phi$. As stated earlier, the parameter α reflects firm's bargaining power in renegotiation with the local governments, who have different political preferences (left-wing or right-wing); and rightists may be more prone to favor private firms than leftists ([37]). To capture such ideological effects, we assume $\alpha(right; \kappa) = \kappa_0 + \kappa_1 right$, where $right$ is a binary variable taking value 1 for right-wing government and 0 otherwise. Nevertheless, recall that α and the cost of public funds λ are not separately identified since only the ratio $\alpha/(1+\lambda)$ appears at the equilibrium conditions in (3.7) and (3.10). The empirical studies suggest that λ is in the interval $[0.15, 0.40]$ in an efficient tax systems ([38]). We choose $\lambda = 0.3$ as in [18].

Define $\omega \equiv (\kappa_0, \kappa_1, \sigma^2, \varphi)$, which includes the parameters in the distribution of innate cost and firms' bargaining power. The equilibrium equation (3.10) provides us with moment conditions that can be used to estimate ω. Specifically,

$$\widehat{\omega} = \underset{\omega}{\operatorname{argmin}} \left(n^{-1} \sum_{i=1}^{n} \varsigma_i \right), \tag{3.38}$$

where n is the sample size, i.e., the total number of CF and FF contracts, and ς_i is defined as:

$$\begin{aligned}\varsigma_i &= \Big\{\left[1-\alpha(right_i;\kappa)(1+\lambda)^{-1}\right]\left[F(\widehat{\theta}_2^*(z_i);\mu(z_i),\sigma^2)-F(\widehat{\theta}_{1i}^*;\mu(\widetilde{z}_i),\sigma^2)\right] \\ &\times \left(2\widehat{\beta}_2\widehat{\gamma}_2\widehat{\theta}_2^*(z_i)+\widehat{\beta}_1\widehat{\gamma}_2+\widehat{\beta}_2\widehat{\gamma}_1\right)-[H(\widehat{\theta}_2^*(z_i),z_i)-\bar{b}_i^R]f(\widehat{\theta}_2^*(z_i);\mu(z_i),\sigma^2)(\widehat{\beta}_2+\widehat{\gamma}_2)\Big\}^2.\end{aligned}$$

It is worth noting that the fixed price $\bar{b}_i^R$ is not reported for FF contracts, and they are estimated by plugging estimates $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\gamma}_1, \hat{\gamma}_2, \hat{\pi}_1, \hat{\pi}_2)$ and the covariates vector z of FF contracts into the first equation of (3.37). Alternatively, we may estimate ω, π_1, π_2, and r together using (3.38). However, our results show that the estimates of such an approach are relatively noisy because more parameters are involved than our multiple-step procedure.

3.4.2.3 Welfare Gains of Commitment

To estimate the welfare gains of contracts under commitment relative to those under renegotiation, we first use the previously estimated parameters to simulate the hypothetical fixed price $b^C(z)$ that would be set by the principal in a long-term fixed-price contracts as specified in proposition 1. This is done by first solving (3.7) to obtain the simulated cut-off type $\widehat{\theta}^*(z)$:

$$\begin{aligned}& 4\left[1-\alpha(right;\widehat{\kappa})(1+\lambda)^{-1}\right]F(\theta^*;\widehat{\mu}(z),\widehat{\sigma}^2)(\widehat{\beta}_2\widehat{\gamma}_1+\widehat{\beta}_1\widehat{\gamma}_2+2\widehat{\beta}_2\widehat{\gamma}_2\theta^*) \\ =\;& [4\widehat{\beta}_2^2\theta^{*2}+4\widehat{\beta}_2(\widehat{\beta}_1-\widehat{\gamma}_1)\theta^*+(\widehat{\beta}_1-\widehat{\gamma}_1)^2]f(\theta^*;\widehat{\mu}(z),\widehat{\sigma}^2).\end{aligned}$$

Next, the fixed price $\widehat{b}^C$ is simulated according to (3.6)

$$\widehat{b}^C=-\frac{(\widehat{\beta}_1-\widehat{\gamma}_1)^2}{4(\widehat{\beta}_2+\widehat{\gamma}_2)}+\frac{\widehat{\beta}_1\widehat{\gamma}_2+\widehat{\beta}_2\widehat{\gamma}_1}{\widehat{\beta}_2+\widehat{\gamma}_2}\widehat{\theta}^*+\frac{\widehat{\beta}_2\widehat{\gamma}_2}{\widehat{\beta}_2+\widehat{\gamma}_2}\widehat{\theta}_i^{*2}+z'\widehat{\beta}_3.$$

Let SW^R and SW^C denote the social welfare of the contracts under renegotiation and commitment, respectively.

$$\begin{aligned} SW^R(z) &= S - (1+\lambda)T^R(z) + \alpha(right;\kappa)U^R(z), \\ SW^C(z) &= S - (1+\lambda)T^C(z) + \alpha(right;\kappa)U^C(z), \end{aligned}$$

where $T^R(z)$ and $T^C(z)$ are subsidy (tax) under renegotiation and commitment, respectively, and $U^R(z)$ and $U^C(z)$ are their informational rent (profit) counterparts. The definition of these objectives are included in the appendix. Consequently, the welfare gains from the commitment is

$$\begin{aligned} \Delta SW(z) &\equiv SW^C(z) - SW^R(z) = \alpha(right;\kappa)\big[U^C(z) - U^R(z)\big] \\ &\quad -(1+\lambda)\big[T^C(z) - T^R(z)\big] \\ &= \alpha(right;\kappa)\Delta U(z) - (1+\lambda)\Delta T(z), \end{aligned} \tag{3.39}$$

where the first and second terms on the R.H.S. of the last line are difference of weighted informational rent and social cost, respectively between two contracts. By integrating out the covariates z, we obtain the average social welfare gain $\overline{\Delta SW} = \int_{z\in\mathbf{Z}} \Delta SW(z)dG(z)$, where $G(z)$ is the cdf of z. To compute the average welfare gain $\overline{\Delta SW}$, we plug the previously estimated and simulated parameters into its definition.

3.4.3 Results

Table 3.2 reports the estimates of parameters for both cost and disutility functions. The estimates suggest that the realized cost is convex in the size of the firms' innate cost, thus providing strong empirical evidence against the linear cost assumed in existing literature. An important implication is that firms enjoy increasing returns to scale in the cost it induces, which provides less efficient firms with incentives to exert more effort. The disutility function is also estimated to be convex, though the coefficient of the quadratic term is not statistically significant. This implies that the marginal disutility of exerting cost-reducing effort is increasing in the level of effort.

Table 3.2: Regression results of cost and disutility

Functions	Variables	(1) CC contracts	(1) FF+CF contracts	(2) CC contracts	(2) FF+CF contracts
8*Cost	Ineff.	9.815***	9.815***	9.361***	9.361***
		(3.444)	(3.444)	(2.579)	(2.579)
	Ineff. $\times$ Ineff.	0.005	0.005	0.010***	0.010***
		(0.004)	(0.004)	(0.003)	(0.003)
	Labor fee	0.892***	0.985***	0.801***	0.946***
		(0.085)	(0.078)	(0.064)	(0.077)
	Network size	6.046***	6.874***	5.531***	6.316***
		(0.769)	(0.979)	(0.586)	(0.953)
	Right	651.7***	-1756.0***	799.1***	-1474.0***
		(112.6)	(308.4)	(87.7)	(309.1)
	Connex			-171.9	-1795.9***
				(122.7)	(421.5)
	Agir			-1038.5***	-1407.4***
				(97.6)	(519.9)
	Tran			134.701	-1601.6***
				(268.6)	(351.6)
2*Disutility	Effort	48.19	48.19	88.75**	88.75**
		(57.89)	(57.89)	(43.37)	(43.37)
	Effort $\times$ Effort	0.003	0.003	0.002	0.002
		(0.041)	(0.041)	(0.004)	(0.004)
N		174	369	174	369

Standard errors in parentheses are bootstrapped 1000 times. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Not surprisingly, our results show that cost increases with labor fee and network size. Specifically, the marginal cost of labor fee is close to unit under both CR and FP schemes. This reinforces that the transport industry is labor intensive and further justifies our usage of employee number as a measurement for innate cost. The marginal cost of network size is similar between FP and CR regimes: one additional kilometer network will incur about 6K cost. If the local government is right-wing the cost will be higher under CR regime, and lower under FP regime. The reason for this opposite effect might be due to the right-wing government's favor of the higher-power incentive of FP contracts over CR contracts. Finally, there is strong empirical evidence that group fixed-effects exist for both CR and FP contracts, which suggests heterogeneous managerial structures of different groups.

The estimated dependence of the two cut-off types on the covariates z is presented in Table 3.3. We observe from the results that two cut-off types are dependent on the characteristics of firms and contracts differently: the first cut-off type θ_1^* is mainly affected by the network size, which varies at the contract-level, whereas the second one θ_2^* relies on both contract and firm-level variables. Recall that firms with $\theta < \theta_1^*$ and $\theta > \theta_2^*$ are most and least efficient segments, respectively. Thus the estimates suggest that the firms in the most efficient segment are similar, whereas those in the least efficient segment are more heterogeneous. We also estimate the intertemporal weight r to be 0.919 with the standard deviation being 0.150, which implies that firms pay most attention to the profit of the first-period.

Table 3.4 shows the estimates of the distribution of innate cost and the bargaining power. The main determinant of the mean of innate cost μ is the group effect. Specifically, Tran group has the largest average innate cost, implying that firms owned by Tran is least efficient on average. Such a result may be due to the fact that Tran is a semi-public group and hence less efficient while others are privately owned and more efficient. The bargaining power for leftists and rightists are estimated to be close to each other (1.261 for leftists versus 1.298 rightists), which indicates that both left-wing and right-wing authorities weigh similarly the firms' profits when evaluating the social welfare.

The results on the welfare comparison between contracts under renegotiation and commitment are as follows. First, our estimate of the subsidy under commitment (T^C, \$

Table 3.3: Estimates of cut-off types

Parameters	θ_1^*	θ_2^*
Labor	-0.055	0.464***
	(0.104)	(0.127)
Network size	0.420***	1.890***
	(0.096)	(0.165)
Connex	33.85	33.86
	(23.80)	(698.23)
Agir	47.50	47.50
	(30.30)	(4418.0)
Tran	6.027	12775.3***
	(22.065)	(4456.8)
Constant	0.304***	1.366***
	(0.103)	(0.149)

Standard errors in parentheses are bootstrapped 1000 times.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

407,000) is about 9 percent less than that under renegotiation (T^R, \$443,000), suggesting that taxpayers are welfare gainers under commitment. Second, the profit of firms under commitment (U^C, \$65,800) in our model is about 60 percent larger than under renegotiation (U^R, \$41.200), implying that firms are also gainers of welfare. This well explains the lobbying effort of firms into extending contract length or equivalently seeking contract offered by the principal under commitment. Third, the total welfare gains by switching from negotiation to commitment (ΔSW) are 78.7 million, which is substantial. More importantly, our results show that sixty percent of the gains ($(1 + \lambda)\Delta T/\Delta SW$) would accrue to taxpayers (the principal) and the firms obtain the remaining forty percent.

Our results reinforce the finding in [18] that switching from contracts under renegotiation to that under commitment leads to substantial welfare gain. Nevertheless, our empirical results suggest that both firms and taxpayers are welfare gainers under commitment while taxpayers would loss about 22 percent of welfare in [18]. Furthermore, we find a much larger total welfare gains (78.7 million) than [18] (2.1 million). The discrepancy of social welfares is mainly attributed to both the theoretical implication by convexity of

Table 3.4: Estimates of the distribution of innate cost and bargaining powers

Parameters		Estimates
Bargaining power for Leftist		1.261***
		(0.404)
Bargaining power for Rightist		1.298*
		(0.735)
S.t.d. of innate cost (σ^2)		265.43*
		(157.27)
12*Mean of innate cost (μ)	Labor	0.948
		(1.178)
	Network size	-2.129
		(5.256)
	Connex	-6.583*
		(3.856)
	Agir	-122.94***
		(5.033)
	Tran	35.62***
		(4.391)
	Constant	-12.07***
		(3.965)

Standard errors in parentheses are bootstrapped 1000 times.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

firms' cost function and the empirical implications by comparisons of cut-off types under two mechanisms. On the one hand, As stated in the theoretical model, when cost function is convex, the marginal benefit of effort is increasing in innate cost. Hence, those inefficient firms (with higher innate costs) have stronger incentives to exert cost-reducing effort if they choose FP contracts. On the other hand, we draw the average cut-off types and firms' choices of contracts for both mechanisms based on our estimates, as is illustrated in Figure 3.1, there is a larger portion of firms choose FP contracts under commitment than under negotiation. Specifically, the figure shows that $\theta_1^* < \theta^* < \theta_2^*$ and $\theta_2^* - \theta^* < \theta^* - \theta_1^*$. Note that under commitment, all the firms with $\theta \in [\underline{\theta}, \theta^*]$ choose FP contracts, whereas under negotiation those firms with $\theta \in [\underline{\theta}, \theta_1^*]$ and $\theta \in [\theta_1^*, \theta_2^*]$ choose FP contracts in both periods and the second period, respectively. Considering the empirical result that that $\theta_2^* - \theta^*$ is much smaller than $\theta^* - \theta_1^*$, it is most likely that if we take both periods into account, there would be more firms exerting cost-reducing effort under commitment, thereby producing more welfare gains by combination with the theoretical implication above. Despite the quantitative difference of welfare gains in this paper and [18], both results are qualitatively identical in the sense that the renegotiation and/or transaction cost associated with contracts under renegotiation is substantial in the public transport procurement in France.

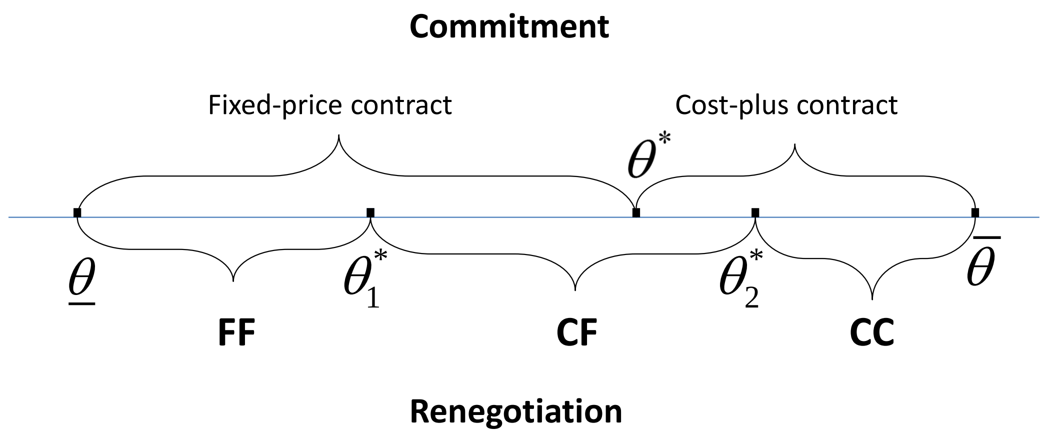

Figure 3.1: Illustration of firms' choice

3.5 Conclusion

We provided a rigorous econometric analysis of the two-period FPCR contracts under both renegotiation and commitment, where we generalize the widely used identity cost function to a convex one and allow for a heterogeneous disutility function of effort. We proved that the model is nonparametrically identified if firms exert effort and the result of identification can be applied to a large class of contracts with and without incentives. If we include the firms without exerting effort, the model is semi nonparametrically identified. Based on the identification argument, we propose a feasible procedure to estimate the model primitives. Using the public transport procurement contracts in France, we found that cost function of firms are convex and the convexity has important implications for the welfare analysis: if the contract is switched from renegotiation to commitment, both taxpayers and firms would benefit and the major gains of social welfare would accrue to taxpayers.

4. NONPARAMETRIC IDENTIFICATION AND ESTIMATION OF CONTRACTING WITH EXTERNALITIES

4.1 Introduction

Due to the fundamental role of asymmetric information in economic relations, during the past decades contracts have flourished and dedicated to how the information asymmetry and incentives induce strategic behavior among economic agents in a number of directions (see [15]; and [16]). One branch of contracts focuses on bilateral contract between one principal and one agent, in which the agent's hidden information and hidden action lead to the adverse selection and moral hazard problems, respectively. For example, [2] propose a contract model to combine adverse selection and moral hazard in procurement and regulation, and [3] establish the nonparametric identification and estimation in an extended model with adverse selection and moral hazard.

The second branch of of bilateral contracts between one principal and multiple agents emphasizes the externalities generated by the dependence of one agent's payoff on other agents' contracts, while the first branch of contracts involves no externalities due to the setting of one principal and one agent. This paper considers the contracting with externalities developed by [7]. In the model, the principal simultaneously makes one offer to each agent, and then agents simultaneously decide whether to accept or reject their own offers. Because those offers are private in the sense that each agent only observes his own offer, it is a dynamic incomplete-information game. In actuality, the model has found wide applications in various economic situations, including vertical contracts in which profits of downstream firms depend on all downstream firms' contracts with the upstream firm ([9]); exclusive dealing in which payoffs of agents rely on the number of agents who sign exclusive contracts with the principal ([10]; [11]); network externalities ([12]; [13]); and among others. Despite the wide use of contracting with externalities in various sectors such as medical device industry and publishing industry, there are few empirical studies on the incomplete-information contracting game with externalities. This paper constitutes the first effort on the rigorous econometric analysis of contracting with externalities.

First, we show the existence of equilibrium in the model under two out-of-equilibrium beliefs, respectively. Due to the essence of incomplete information implied by the privacy of contracts, different out-of-equilibrium beliefs often induce different equilibrium outcomes. Although arbitrary beliefs can be assigned following the out-of-equilibrium offer, there are two widely used out-of-equilibrium beliefs (i.e., passive beliefs and symmetric beliefs) in the theoretical literature. When receiving an out-of-equilibrium offer, with passive beliefs the agent believes that other agents still receive equilibrium offers, while with symmetric beliefs the agent adjusts his initial beliefs about other agents' offers by the same deviations. As expected, the equilibrium outcomes are different between passive and symmetric beliefs.

Next, we establish the nonparametric identification of structural elements, including the principal's cost function, the agent's payoff function, bargaining power, and the joint distribution of agents' shocks, under passive beliefs and symmetric beliefs, respectively. Nonparametric identification is valuable in terms of providing which information in the data allows identification of each unknown function and avoiding potential parametric misspecifications. Given the observed quantity and payment in the contract, the nonparametric identification strategy relies on the one-to-one mapping between the quantity vector of contracts and the shock vector of agents. The one-to-one mapping is key to identifying incomplete-information games such as auctions in [34], in which the bidder's private value and his bid are analogous to the agent's shock and the quantity in the contract. Besides the unobservability of the agent's shock, the main difficulty for identification arises from the multiple unknown functions given only two observed variables in the contract. For the goal of identification, we will explore mild restrictions on the joint density of the shock vector, the payoff function, and the variation of agents' characteristics.

In addition, we propose a multi-step nonparametric estimation procedure and establish the rates of uniform convergence of nonparametric estimators, including sieve estimators and kernel estimators. The convergence rate relies on the recent literature on sieve estimators. [39] propose sieve estimators of multiple functions by using conditional moment conditions, and establish the rates of uniform convergence under regular conditions. Our convergence results build on [39] by providing primitive sufficient conditions.

Lastly, we apply the model to coronary stents contracts between the manufacturer and hospitals in the United States, and find that passive beliefs fit the dataset better than symmetric beliefs. More relevantly, the counterfactual result shows that if the bargaining power of hospitals increases, under passive beliefs the stent's price decreases by some reasonable amount, while under symmetric beliefs the decrease of price is unreasonable.

4.1.1 Preview of the Model, Identification, and Estimation

The model builds on the seminal work of [7] by allowing for a positive bargaining power of agents in the determination of payment in the offer, instead of an extremely zero bargaining power of agents. The contracting game proceeds as follows. In the beginning, each agent's shock is jointly drawn by Nature and is common knowledge to all players. Next, the principal simultaneously makes an offer of quantity to each agent, and each agent has some beliefs about other agents' offers of quantities according to his offer. Based on his beliefs, both the principal and the agent expect that they use Nash bargaining solution to determine the corresponding monetary payment from the agent to the principal. The payment for each agent is simultaneously determined through Nash bargaining solution. Due to the incomplete-information implied by the fact that each agent can not observe any other agents' offers, this paper provides the existence of the perfect Bayesian equilibrium in the contracting game with two widely used out-of-equilibrium beliefs–passive beliefs and symmetric beliefs, though [7] considers only passive beliefs under the extreme bargaining assumption. When receiving an out-of-equilibrium outcome, with passive beliefs the agent does not update his initial belief and believes that other agents still receive equilibrium offers, while with symmetric beliefs the agent adjusts his initial beliefs about other agents' offers by the same deviation of quantity in the offer. We find that the equilibrium outcomes are different between passive beliefs and symmetric beliefs.

The identification of contracting with externalities proceeds the estimation. We show that the equilibrium conditions implied by the model, together with some functional restrictions on the joint density of shocks, the payoff function, and the variation of agents' characteristics, enable us to nonparametrically identify the model primitives from the joint distribution of quantities and payments in the offers. The model primitives include the principal's cost function, the agent's payoff function, bargaining power, and the joint distribution of agents' shocks. The identification argument under passive beliefs is readily carried over to symmetric beliefs.

Our identification argument takes several steps. First, we provide mild restrictions on the joint density of the shock vector, the payoff function, and the variation of agents' characteristics to identify the payoff function. The identification builds on the influential work [40] who provides a general identification result on the identification of nonparametric simultaneous equations. In order to apply her result to the present model, we explore the one-to-one structural link between the shock vector and the quantity vector, and the rank condition on the matrix involving the payoff function and the joint distribution of shocks. Second, we use the result in [40] to identify another one-to-one mapping between the aggregate shock and the aggregate quantity. Then by utilizing the variation of agents' characteristics, we recover the cost function. Based on these two identified objects, we can back out each agent's shock according to the equilibrium condition, and hence identify the joint distribution of the shock vector. Third, we use the payment equation to identify the bargaining power since it can be expressed as a function of all the identified objects above.

Given the positive identification result, we propose nonparametric estimators of model primitives and establish the rates of uniform convergence of these estimators under regular conditions. First, we use the sieve minimum distance method proposed by [41] to estimate the cost function, bargaining power function, and the first derivative of payoff function according to the conditional moment condition in the payment equation. To allow for the potential correlation between the agent's shock and the payment error, we use the instrument variable to construct the conditional moment condition for the sieve estimation. Following [39], we provide primitive conditions to establish the rates of uniform convergence of those sieve estimators. Second, we use the kernel estimator to estimate the remaining part of the payoff function and show its rate of of uniform convergence. Third, we propose the kernel estimator of the joint distribution of the shock vector by using the pseudo shock vector. The prior estimation results imply the rate of uniform convergence of the pseudo shock vector. By combining the uniform convergence result on the kernel estimator of the multivariate distribution with the true shock vector, we establish the rate of uniform convergence of the kernel estimator of the shock vector's distribution.

4.1.2 Preview of Empirical Findings

We apply our model to study the coronary stents contracts between manufacturers and hospitals in the United States. The objective of the empirical study is to test which belief fits the dataset better and compare the counterfactual price effect of bargaining power between two beliefs. Although typical linear or nonlinear specifications are consistent with conditions for nonparametric identification, our empirical analysis is based on a parametric approach since the test method applies to parametric specifications and the sample size in our dataset is limited.

Specifically, we first estimate the model parameters and then based on the estimates, we use the nonnested test method to test for the asymptotic equivalence of two belief systems. The test result rejects the null hypothesis of asymptotic equivalence in favor of the alternative hypothesis that the model with passive beliefs fits the data better than symmetric beliefs. More relevantly, the counterfactual result shows that if the agent's bargaining power increases, under passive beliefs the stent's price decreases by some reasonable amount, while under symmetric beliefs the decrease of price is unreasonable.

4.1.3 Relation to Existing Literature and Contributions

First of all, this paper is related to identification of nonparametric functions with nonseparable errors. A growing literature on the identification of a system of simultaneous nonseparable equations has attracted more attention in the past decade. [40] provides positive identification results for a fully nonparametric simultaneous model. Based on this identification result, [42] further shows the identification of the simultaneous equations with exogenous regressors without requiring large support conditions on the observable exogenous regressors. For the particular class of triangular models with nonseparable errors, recent literature includes [43], [44], [45], [46], and among others. Besides, it is also related to the identification of the single nonseparable function. For example, [47] proposes various normalizations to identify a nonseparable nonparametric function with monotonicity, while [48] establish that in the absence of monotonicity, the quantiles identify local average structural derivatives of nonseparable models.

The main contribution of the paper is to provide sufficient conditions to nonparametrically identify contracting with externalities. First, based on the identification result in [40], the paper imposes mild functional restrictions on the joint density of the shock vector, the payoff function, and the variation of agents' characteristics, to identify the agent's payoff function. By exploring the one-to-one mapping between the shock vector and the quantity vector, we obtain a similar result as in [40]. Then, we provide sufficient conditions on the joint density of the shock vector, the large support of agents' characteristics, and the additively separability of payoff functions to identify the first part of payoff function under . Those first two conditions are mild and adopted by [40] for the identification of nonparametric simultaneous equations model with a linear index structure. Instead of fully identifying the functions themselves, however, [40] establishes the identification of some features of the functions in terms of a rank condition. Hence, we show the identification of the first part of payoff functions by using the additively separability of payoff functions. Besides, we use [40] to identify a one-to-one mapping between the aggregate quantity between the aggregate shock and hence the second part of payoff functions. Thereby, we recover the cost function. To our best knowledge, this paper is the first to apply the result in [40] to the identification of economic theories.

Second, this paper contributes to the uniform convergence of nonparametric estimators of model primitives in games. One typical game is auctions because the bidder's value in auctions is analogous to the agent's shock in contracting with externalities. [34] establish the rate of uniform convergence of kernel estimators of distribution of independent values in auctions, and [49] extend the uniform convergence result in auctions with affiliated private values. This paper also establishes the rate of uniform convergence of nonparametric estimator of joint distribution of agents' shocks. However, it is more challenging to obtain the uniform convergence in the present paper because it is built upon the convergence rate of sieve estimator of unknown cost function and payoff function (e.g., [41]; [39]), while the rate in the literature on auctions relies only on the convergence rate of kernel estimator of the known distribution of observed bids. Another related game is bargaining models. [50] propose consistent estimators in a stochastic sequential bargaining model. We not only provide the consistent estimation in contracting game with externalities, more importantly, we establish the rate of uniform convergence because it is crucial for recovering the shape of model primitives ([34]).

Third, this paper contributes to a fast-growing empirical analysis on various contracting circumstances with externalities. Although the theoretical literature has been very active, there are few empirical analysis of contract models. This may arise from he difficulty of accessing data on contracts because many contracts are private. During the past decades it is more likely for researchers to obtain data on contracts, and thereby, the empirics on contracts has made much progress. For example, in a mall where anchor stores (department stores) generate positive externalities by drawing customer traffic to other stores, [51] show that mall rental contracts are designed to efficiently price the net externality of each store. [52] use the dataset on magazine distribution to analyze how an upstream firm (publisher) determines the size of its distribution network of downstream firms (retail outlets). [53] evaluates the foreclosure effect of exclusive distribution arranges on competition in the Chicago beer market in 1994. However, the existing empirical literature does not explicitly refer to the incomplete information in contracting with externalities. To our knowledge, this paper is the first to empirically analyze the dynamic incomplete-information contracting with externalities for which the specification of out-of-equilibrium beliefs affects equilibrium outcomes and hence empirical resutls. Using data on contracts in the coronary stent industry, we estimate the model and and find that passive beliefs explain the dataset better than symmetric beliefs. Moreover, the counterfactual analysis shows that the price effect of bargaining powers is more reasonable under passive beliefs than under symmetric beliefs, suggesting the empirical relevance of passive beliefs in the stent industry in the United States.

4.1.4 Roadmap

The rest of the paper is as follows. Section 2 presents the model of contracting with externalities. Section 3 showes the nonparametric identification. Section 4 proposes nonparametric estimators and establishes their rates of uniform convergence. Section 5 is the empirical application. Section 6 concludes. Proofs, figures and tables are collected in the appendix.

4.2 The Model

This paper considers a two-stage incomplete information contracting game in which a risk-neutral principal wishes to trade with $N \geq 2$ risk-neutral agents.[1] The agents's shocks of payoff are drawn by the nature, and are common knowledge. In the first stage the principal simultaneously makes one quantity offer q_i to each agent $i \in N$ (with a slight abuse of notation, N will represent the set as well as the number of agents), where $q_i \in \mathcal{Q}$ is the quantity of trade, $\mathcal{Q}$ is a compact subset of the set $\mathbf{R}_+$ of nonnegative real numbers. In the second stage, agent i has some beliefs about other agents' offers of quantities according to his quantity offer. Based on his beliefs, both the principal and agent i expect that they use Nash bargaining solution to determine the corresponding monetary payment $p_i \in \mathbf{R}$ from agent i to the principal. The payment for each agent is simultaneously determined through Nash bargaining solution. The contracting game is incomplete information due to the privacy of offers $\{(q_i, p_i)\}_{i=1}^N$ in the sense that each agent only observes his own offer.

Externalities among agents arise whenever each agent's payoff depends on other agents' trades. This paper considers a particular class of contracting with externalities. The payoff of agent i is assumed to be $q_i[u(x) + \epsilon_i] - p_i$, where $x \equiv \sum_{i \in N} q_i$ is the aggregate quantity of the trade vector $\boldsymbol{q} = (q_1, \cdots, q_N)'$, $u(\cdot)$ can be interpreted as the deterministic marginal revenue function. The random variable ϵ_i captures agent i's unobserved factors (by researchers) which affect his payoff, such as demand shock or unobserved characteristics of agent i (e.g., see [54]; [55]; and [56]). For ease of exposition hereafter we call ϵ_i as shock. The shock vector $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_N)'$ is assumed to be drawn from a cumulative distribu-

[1] We shall use feminine pronouns for the principal and masculine ones for the agent.

tion function $F_{\boldsymbol{\epsilon}}$ prior to that the principal makes simultaneous offers, and it is common knowledge to all parties but not observed by researchers.[2] The payoff of the principal is assumed to be $y - c(x)$, where $y \equiv \sum_{i=1}^{N} p_i$ is the aggregate payment of the payment vector $\boldsymbol{p} = (p_1, \cdots, p_N)'$, and $c(\cdot)$ is the principal's cost function.

The specification of payoff function is general in theory, there exists empirical evidence from the related literature. As for the theory, this specification includes both the vertical contracting games studied by [9] and the insurance models with moral hazard studied by [57] and [58] as special cases. As for the empirics, the reimbursement rate (the major source of hospital's profit) from patient's insurer to hospital depends on the aggregate demand of hospitals in the United States ([59]; [60]). For the goal of investigating timing incentives of commercial radio stations, [61] specifies the payoff of one station to depend on the proportion of other stations who choose the same action. For the inference of interaction effects in discrete simultaneous games, [62] specify the payoff of each player to rely on the aggregation of all other players' actions.

The specification of out-of-equilibrium beliefs is critical to perfect Bayesian equilibria (PBE) of incomplete-information games since arbitrary beliefs can be assigned following the principal's out-of-equilibrium offers, which gives rise to different equilibrium outcomes. To consider the principal's incentive to deviate from an equilibrium outcome $\{(q_i^*, p_i^*)\}_{i=1}^{N}$, according to the existing literature, there are two widely-used out-of-equilibrium beliefs: passive beliefs and symmetric beliefs.[3] When agent i receives an out-of-equilibrium offer $q_i \neq q_i^*$, passive beliefs imply that he believes that other agents still receive their equilibrium offers $\{q_j^*\}_{j \neq i}$ and pay the corresponding equilibrium pay-

[2]Although we assume the common knowledge of $\boldsymbol{\epsilon}$, this game is incomplete-information due to the privacy of offers. Allowing for private random shocks (i.e., ϵ_i is known only to agent i) introduces the second source of incomplete information and complicates the theoretical analysis, which beyonds the scope of this paper.

[3]Here we can not use Bayesian rule to update belief on out-of-equilibrium paths because there is no prior incomplete information and hence no prior belief.

ments $\{p_j^*\}_{j\neq i}$, while symmetric beliefs imply that he believes that other agents receive the same deviation offers. Since there is neither general agreement nor empirical evidence on which belief is more appropriate, we will analyze the equilibrium of contracting with externalities under passive beliefs and symmetric beliefs, respectively.[4]

4.2.1 Passive Beliefs

Passive beliefs have been intensively employed in the contracting game with one principal and multiple agents, including [63], [9], [7], [64], and among others. When agent i receives an out-of-equilibrium offer $q_i \neq q_i^*$, under passive beliefs he does not expect multilateral deviations and continues to believe that all other agents still receive their equilibrium offers $\{q_j^*\}_{j\neq i}$ associated with the corresponding equilibrium payments $\{p_j^*\}_{j\neq i}$. As [65] state, one rationale for passive beliefs is that in such a Cournot-like context, the quantity actually sold to one agent does not directly affect the payoff the principal derives from other contracts.

We start with the determination of payment in the out-of-equilibrium offer (q_i, p_i). [7] assumes that the principal has all the bargaining power, and hence $p_i = q_i\big[u\big(q_i + \sum_{j\neq i} q_j^*\big) + \epsilon_i\big]$, which implies that the agent's payoff is zero. However, agents may have positive bargaining power, of which the source includes asymmetric preferences of the agents and the principal, their different beliefs regarding the possibility of a breakdown, and some determinants of economic environments ([66]; [58]), and hence the agent's payoff is positive. In the vertical contracting literature some papers analyze the equilibrium when downstream firms have some bargaining power (e.g., [67]). Indeed, [63] adopt the Nash bargaining solution in bilateral negotiations with one upstream firm and two downstream firms by allowing for symmetric bargaining powers between two parties. Due to

[4] In the related literature, the third type of beliefs is wary beliefs proposed by [9]. However, the solution of wary beliefs is theoretically challenging in our model due to the nonparametric form of structural functions, and we do not consider wary beliefs in the paper.

the theoretical attractions, Nash bargaining solution has been widely used in empirical analysis on bilateral negotiations with externalities (e.g., [68]; and [69]).[5] To make more realistic predictions in the sense that the agent's payoff is positive, this paper also nests the Nash bargaining solution in the contracting game to determine the payment in the deviation offer q_i given the equilibrium contracts $\{(q_i^*, p_i^*)\}_{i=1}^N$ and q_i.[6] Note that here the Nash bargaining solution applies because it is solved given the *believed* offers $\{(q_i^*, p_i^*)\}_{i=1}^N$ under passive beliefs and the deviation trade q_i, though the model as a whole is incomplete information.

Specifically, given the equilibrium contracts $\{(q_j^*, p_j^*)\}_{j\neq i}$ and the deviation trade q_i, the Nash bargaining solution for p_i is given by

$$p_i \;=\; \operatorname*{argmax}_{\widetilde{p}_i \geq 0} \left\{ q_i \left[u\Big(q_i + \sum_{j\neq i} q_j^*\Big) + \epsilon_i \right] - \widetilde{p}_i \right\}^{\gamma} \left\{ \widetilde{p}_i + \sum_{j\neq i}^{N} p_j^* - c\Big(q_i + \sum_{j\neq i} q_j^*\Big) \right\}^{1-\gamma},$$

together with the disagreement point $(0,0)$, where $\gamma \in (0,1)$ represents the bargaining power of agent i in negotiation with the principal. Then, the Nash bargaining solution for p_i is

$$p_i = \gamma c\Big(q_i + \sum_{j\neq i} q_j^*\Big) + (1-\gamma) q_i \left[u\Big(q_i + \sum_{j\neq i} q_j^*\Big) + \epsilon_i \right] - \gamma \sum_{j\neq i}^{N} p_j^*.$$

Note that $p_i = q_i\big[u\big(q_i + \sum_{j\neq i} q_j^*\big) + \epsilon_i\big]$ when $\gamma = 0$, which implies that the general Nash bargaining solution includes [7] as a special case.[7] Thus, the principal's payoff from

[5]This solution abstracts away the details of the bargaining process and considers only the set of outcomes that satisfy certain 'reasonable' axioms. [70] and [66] show that the Nash bargaining solution in a bilateral setting corresponds to the unique subgame perfect equilibrium of a dynamic complete-information alternating offers game. [63] nests a Nash bargaining solution within a Nash equilibrium without a complete non-cooperative structure.

[6]See the use of the Nash bargaining solution in related contracting games with private negotiations in [67].

[7]Although the results on Nash bargaining solution in [70] and [66] apply to complete-information games, the relevance of Nash bargaining solution in the model is that given agent i's beliefs, both the principal and agent i can consider the subgame as complete-information. Therefore, they may participate in the traditional complete-information alternating offers game to determine the payment given agent i's beliefs about other agents' quantities.

making deviations $\boldsymbol{q}$ is

$$\pi_1(\boldsymbol{q}, \boldsymbol{q}^*, \gamma|\boldsymbol{\epsilon}) \equiv \sum_{i=1}^{N} \left\{ \gamma c\left(q_i + \sum_{j \neq i} q_j^* \right) + (1-\gamma) q_i \left[u\left(q_i + \sum_{j \neq i} q_j^* \right) + \epsilon_i \right] \right\} - c\left(\sum_{i=1}^{N} q_i \right).$$

The equilibrium concept implies that the principal has no incentive to deviate from $\{(q_i^*, p_i^*)\}_{i=1}^N$, that is, $\boldsymbol{q}^* \in \underset{\boldsymbol{q} \in \mathcal{Q} \times \cdots \times \mathcal{Q}}{\text{argmax}} \ \pi_1(\boldsymbol{q}, \boldsymbol{q}^*, \gamma|\boldsymbol{\epsilon})$. Thus, the Perfect Bayesian Equilibrium can be characterized below.

Definition 1. *A strategy profile $(\boldsymbol{q}^*(\boldsymbol{\epsilon}), \boldsymbol{p}^*(\boldsymbol{\epsilon})) \equiv \{(q_i^*(\boldsymbol{\epsilon}), p_i^*(\boldsymbol{\epsilon}))\}_{i=1}^N$ constitutes a Perfect Bayesian Equilibrium under passive beliefs if and only if for any $\boldsymbol{\epsilon}$,*

$$\boldsymbol{q}^*(\boldsymbol{\epsilon}) \in \underset{\boldsymbol{q}(\boldsymbol{\epsilon}) \in \mathcal{Q} \times \cdots \times \mathcal{Q}}{\textit{argmax}} \ \pi_1(\boldsymbol{q}(\boldsymbol{\epsilon}), \boldsymbol{q}^*(\boldsymbol{\epsilon}), \gamma|\boldsymbol{\epsilon}).$$

The following lemma provides a set of sufficient conditions for the unique existence of equilibrium of contracting with externalities under passive beliefs.

Lemma 6. ***(uniqueness of equilibrium under passive beliefs****) Suppose that (a) $\boldsymbol{\epsilon}$ is distributed according to $F_{\boldsymbol{\epsilon}}$ with a continuously differentiable density $f_{\boldsymbol{\epsilon}}$ on its support $\mathcal{E}^N \equiv [\underline{e}, \overline{e}]^N$, (b) $(c, u) \in \mathcal{H}_1$, where $\mathcal{H}_1$ denote the set of functions (c, u) such that (i) $u \geq 0$ and $c \geq 0$; (ii) both u and c are triple continuously differentiable and $u'(\cdot) > 0$; (iii) $\pi_1(\boldsymbol{q}, \boldsymbol{q}^*, \gamma|\epsilon)$ is concave in $\boldsymbol{q}$ for any $\boldsymbol{\epsilon}$. Then, there exists a unique Perfect Bayesian Equilibrium $(\boldsymbol{q}^*(\boldsymbol{\epsilon}), \boldsymbol{p}^*(\boldsymbol{\epsilon})) \equiv \{(q_i^*(\boldsymbol{\epsilon}), p_i^*(\boldsymbol{\epsilon}))\}_{i=1}^N$ under passive beliefs.*

Condition (a) imposes standard smoothing conditions on the distribution of $\boldsymbol{\epsilon}$ and functions (u, c). The compactness of the support of $\boldsymbol{\epsilon}$ is required by the compactness of $\boldsymbol{Q}$ used for the existence of equilibrium whenever there exists some regular one-to-one mapping between $\boldsymbol{\epsilon}$ and $\boldsymbol{Q}$ (One exception is the trigonometric function such as tangent. However, this class of one-to-one mappings requires a heuristic support such as $[-\pi/2, \pi/2]$, which is usually implausible for most datasets). Condition (b)-(i) is implicitly assumed in the model, condition (b)-(ii) impose standard smoothing conditions on the functions u and c, and condition (b)-(iii) is an important condition in the fixed point theorem used for the existence of PBE, and it tends to hold under standard primitive conditions, e.g., $c'' > 0$ and $u'' < 0$.

Proof of Lemma 6 As in [7], the existence of equilibrium immediately follows Kakutani's fixed point theorem under the conditions in Lemma 6. Next, we prove the uniqueness of the equilibrium by showing that there is a one-to-one mapping between the optimal quantity vector and the shock vector. To do so, we first show the one-to-one mapping between the aggregate trade x^* and aggregate shock $\xi \equiv \sum_{i=1}^{N} \epsilon_i$. Specifically, under the conditions on differentiability of u and c, the first-order condition and the second order condition are $\epsilon_i = c'(x^*) - u(x^*) - q_i^* u'(x^*)$ and $Nc''(x^*) - 2Nu'(x^*) - x^* u''(x^*) > 0$, respectively, for any $i \in N$. Hence, $\xi = Nc'(x^*) - Nu(x^*) - x^* u'(x^*) \equiv s(x^*)$ and $s'(x^*) = Nc''(x^*) - 2Nu'(x^*) - x^* u''(x^*) + (N-1)u'(x^*) > (N-1)u'(x^*) > 0$. Combining the above results implies the unique $(\boldsymbol{q}^*, \boldsymbol{p}^*)$ for a given $\boldsymbol{\epsilon}$.

4.2.2 Symmetric Beliefs

In contrast to passive beliefs, symmetric beliefs mean that when an agent receives an out-of-equilibrium offer, he believes that other agents receive the same offer. Symmetric beliefs have been widely employed in the private contracting game such as [9], and in related contracting games such as [71] and [72]. As [72] argue, one rationale for symmetric beliefs is that the agent believes that the principal has the incentive to make the same deviation on other contracts.

We redefine symmetric beliefs by allowing for the heterogeneities of agents represented by their respective shocks, while those two justifications for the original symmetric beliefs are plausible under the assumption that agents are identical. In practice, it is common that different agents receive different contracts. Specifically, when agent i receives an out-of-equilibrium offer q_i, he believes that the quantity agent j receives is $B(q_i, q_i^*, q_j^*)$, where $B(q_i, q_i^*, q_j^*) = q_j^* + (q_i - q_i^*)$. The first term q_j^* captures the heterogeneity of agents because equilibrium quantities depend on shocks; the second term $(q_i - q_i^*)$ follows the rationale that agents interpret relative deviations as trembles by the principal and assume that those trembles are perfectly correlated, that is, agent i believes that the quantities received by other agents change by the same relative deviation. In short, with symmetric beliefs agent i revises his initial belief about other agent's offer by adding the same relative deviation $(q_i - q_i^*)$.

Our symmetric beliefs does not only satisfy the consistency requirement of the PBE, it is also important for the well-defined Nash bargaining solution with symmetric beliefs because the Nash bargaining solution with symmetric beliefs is more complicated than passive beliefs. Under passive beliefs p_i is obtained by solving the single equation given the equilibrium contracts $\{(q_j^*, p_j^*)\}_{j \neq i}$ and the deviation trade q_i. However, for the calculation of p_i with symmetric beliefs it is necessary to specify the way in which agent i's belief about other agent j's payment is determined. To this end, we assume that both the principal and agent i anticipate that other agent j's payment is also determined through Nash bargaining solution (see [63]), and that agent i believes that agent j reasons the same way. The latter assumption implies that agent i's belief about agent k's trade given agent i's *received* trade q_i should be equal to agent j's belief about agent k's trade given agent j's *believed* trade $B(q_i, q_i^*, q_j^*)$, that is, $B(q_i, q_i^*, q_k^*) = B(B(q_i, q_i^*, q_j^*), q_j^*, q_k^*)$, which holds with our specification of symmetric beliefs.[8]

Under the above assumptions we can obtain the Nash bargaining solution of p_i as well as agent i's belief about other agents' payments by solving a well-defined system of N

[8]However, the polynomial specification, $B(q_i, q_i^*, q_j^*) = q_j^* + \sum_{k=1}^{K} \lambda_k (q_i - q_i^*)$, does not satisfy this implication unless $\lambda_k = 0$ for all $k = 1, \cdots, K$.

simultaneous payment equations. Let

$$\boldsymbol{q}^i \equiv [B(q_i, q_i^*, q_1^*), \cdots, B(q_i, q_i^*, q_{i-1}^*), q_i, B(q_i, q_i^*, q_{i+1}^*), \cdots, B(q_i, q_i^*, x_N^*)]',$$

be agent i's believed trade vector. Then, the Nash bargaining solution of p_i is

$$\begin{aligned} p_i &= \frac{1-2\gamma+N\gamma}{(1-\gamma)(1-\gamma+N\gamma)}\left[\gamma c(x^i) + (1-\gamma)q_i(u(x^i)+\epsilon_i)\right] \\ &- \frac{\gamma}{(1-\gamma)(1-\gamma+N\gamma)}\sum_{j\neq i}\left\{\gamma c(x^i) + (1-\gamma)[q_j^* + (q_i - q_i^*)][u(x^i)+\epsilon_j]\right\}, \end{aligned}$$

where $x^i = \sum_{j=1}^N \boldsymbol{q}_j^i = x^* + N(q_i - q_i^*)$, and $x^* \equiv \sum_{i=1}^N q_i^*$. Then, $\{(q_i^*, p_i^*)\}_{i=1}^N$ can be sustained in equilibrium under symmetric beliefs if and only if

$$\boldsymbol{q}^* \in \underset{\boldsymbol{q}\in\mathcal{Q}\times\cdots\times\mathcal{Q}}{\text{argmax}}\ \pi_2(\boldsymbol{q}, \boldsymbol{q}^*, \gamma|\boldsymbol{\epsilon}),$$

where

$$\begin{aligned} &\pi_2(\boldsymbol{q}, \boldsymbol{q}^*, \gamma|\boldsymbol{\epsilon}) \\ &\equiv \frac{1}{(1-\gamma)(1-\gamma+N\gamma)}\sum_{i=1}^N \Bigg\{(1-2\gamma+N\gamma)\left[\gamma c(x^i) + (1-\gamma)q_i(u(x^i)+\epsilon_i)\right] \\ &-\gamma\sum_{j\neq i}\left[\gamma c(x^i) + (1-\gamma)[q_j^* + (q_i - q_i^*)](u(x^i)+\epsilon_j)\right]\Bigg\} - c\left(\sum_{i=1}^N q_i\right). \end{aligned}$$

Under similar conditions with Lemma 6, the lemma below provides a set of sufficient conditions for the unique existence of equilibrium of the contracting with externalities under symmetric beliefs.

Lemma 7. *(uniqueness of equilibrium under symmetric beliefs) Under the conditions (a) in Lemma 6, (b)* $(c, u) \in \mathcal{H}_2$*, where* $\mathcal{H}_2$ *denote the set of functions* (c, u) *such that* (i)

$u \geq 0$ and $c \geq 0$; (ii) both u and c are triple continuously differentiable, $u'(\cdot) > 0$, and $c''(\cdot) > 0$; (iii) $\pi_2(\boldsymbol{q}, \boldsymbol{q}^*, \gamma|\epsilon)$ is concave in $\boldsymbol{q}$ for any $\boldsymbol{\epsilon}$. Then, there exists a unique Perfect Bayesian Equilibrium $(\boldsymbol{q}^*(\boldsymbol{\epsilon}), \boldsymbol{p}^*(\boldsymbol{\epsilon})) \equiv \{(q_i^*(\boldsymbol{\epsilon}), p_i^*(\boldsymbol{\epsilon}))\}_{i=1}^N$ under symmetric beliefs.

Proof of Lemma 7 The proof is the similar to that of Lemma 6 with exception that we need the additional condition that $c''(\cdot) > 0$ to guarantee that there is a one-to-one mapping between the aggregate trade x^* and aggregate shock $\xi \equiv \sum_{i=1}^N \epsilon_i$ by combining the first order condition $\epsilon_i = c'(x^*) - u(x^*) - Nq_i^* u'(x^*)$ and the second order condition $[N(1-\gamma)]^{-1}[1-\gamma(N^2-N+1)]c''(x^*) - 2u'(x^*) - x^* u''(x^*) > 0$.

By comparison it is clear that the equilibrium outcome under passive beliefs differs from that under symmetric beliefs. Suppose that both u and c are first continuously differentiable, the first-order condition with respect to the optimal trade vector under passive beliefs is $\epsilon_i = c'(x^*) - u(x^*) - q_i^* u'(x^*)$ for each $i \in N$, while under symmetric beliefs it is $\epsilon_i = c'(x^*) - u(x^*) - Nq_i^* u'(x^*)$. Accordingly, the optimal payment vector differs between passive beliefs and symmetric beliefs. Similar results on the dependence of equilibrium outcomes on passive beliefs or symmetric beliefs can be found in other papers. For example, in a contracting game consisting of a monopoly manufacturer and two independent and competing retailers, [9] show that the manufacturerŠ's profit is higher with symmetry beliefs than with passive beliefs.

4.3 Identification

This section presents the identification of the contracting with externalities under passive beliefs and symmetric beliefs, respectively. Hereafter, we use capital letters to indicate random variables and small letters for their realizations. Denote by $\{(Q_i, P_i)\}_{i=1}^N$ the observed contracts generated under the same equilibrium. Denote by $\widetilde{Z}$ the homogeneous characteristics of agents. Since $\widetilde{Z}$ may vary across observations, our identification arguments should be interpreted as conditional on $\widetilde{Z}$, and hence the model primitives conditional on $\widetilde{Z} = \tilde{z}$ are $\mathcal{S}_{\tilde{z}} \equiv [u(\cdot, \tilde{z}), c(\cdot, \tilde{z}), \gamma(\tilde{z}), F_{\boldsymbol{\epsilon}|\widetilde{Z}=\tilde{z}}(\cdot)]$. In what follows, for a generic multivariate function $a(\cdot)$, we denote the first and second derivative with resect to its first component by $a'(\cdot)$ and $a''(\cdot)$, respectively, and denote the inverse function of $a(\cdot)$ with respect to its first component by $a^{-1}(\cdot)$. Let $\widetilde{Z} \equiv (Z, Z^0)$ with the support $\widetilde{\mathcal{Z}} = (\mathcal{Z}, \mathcal{Z}^0) \subset \mathcal{R}^{d_{\tilde{z}}}$. Since our identification argument holds regardless of the dimension of Z, for simplicity of exposition, we assume that $d_z = 1$ in this section. And, $\mathcal{Z} = [\underline{z}, \overline{z}]$ with $-\infty < \underline{z} < \overline{z} < \infty$. Now we make the following assumption.

Assumption 5. *$\boldsymbol{\epsilon}$ is independent of Z, conditional on Z^0.*

Assumption 5 guarantees that, conditional on Z^0, the distribution of $\boldsymbol{\epsilon}$ is the the same for any value of Z. The conditional independence implies that it allows for the dependence between $\boldsymbol{\epsilon}$ and Z, and also requires that Z^0 involves richer $\boldsymbol{\epsilon}$-related information than Z. Note that the any primitive function in $\mathcal{S}_{\tilde{z}}$ is not necessarily a function of the vector z^0, though we explicitly write z^0 as an argument of these functions. Since the following assumptions and identification results can be interpreted as conditional on Z^0, the inclusion of Z^0 does not add any complication to our identification. Hence, for ease of exposition, we will omit Z^0 from the model, and rewrite the model primitives conditional on $Z = z$ as $\mathcal{S}_z \equiv [u(\cdot, z), c(\cdot, z), \gamma(z), F_{\boldsymbol{\epsilon}}(\cdot)]$. And, we assume both u and c are triple continuously differentiable on its support $\mathcal{X} \times \mathcal{Z}$, and the density $f(z) > 0$ for any $z \in \mathcal{Z}$.

4.3.1 Identification of $u(\cdot)$

As shown in the proof of Lemma 6, there is a one-to-one mapping between aggregate trade $X \equiv \sum_{i=1}^N Q_i$ and aggregate shock $\xi \equiv \sum_{i=1}^N \epsilon_i$,

$$\xi = Nc'(X, z) - Nu(X, z) - Xu'(X, z) \equiv s(X, z), \tag{4.1}$$

where $s'(X, z) > 0$. Hence, $X = s^{-1}(\xi, z)$ and $s^{-1\prime}(\cdot, z) > 0$.

We focus on the identification the function $u(\cdot, z)$ by recovering some features of a system of simultaneous equations. Substituting $s(X, Z)$ into (4.1) gives rise to the system of simultaneous equations

$$\boldsymbol{\epsilon} \equiv \boldsymbol{m}(\boldsymbol{Q}, Z) = (m^1(\boldsymbol{Q}, Z), \cdots, m^N(\boldsymbol{Q}, Z))', \tag{4.2}$$

where $\epsilon_i \equiv m^i(\boldsymbol{Q}, Z) = N^{-1}s(X, Z) + N^{-1}Xu'(X, Z) - Q_i u'(X, Z)$. We will provide a lemma which shows the identification of features of the above simultaneous system, which is important to establish the identification of $u(\cdot)$, since the primitive function $u(\cdot)$ can be viewed as a feature of the simultaneous system $\boldsymbol{m}(\cdot)$. This lemma is built upon [40, 42] who provide the nonparametric identification of features of a system of simultaneous equations with nonadditive unobserved random errors, though not focusing on the identification of the system itself. The identification is defined in terms of the definition of observational equivalence. Roughly, the main idea is to explore the restrictions that independence between X and Z_2 imposes on any function $\widetilde{\boldsymbol{m}}$, and further, those restrictions can be expressed in terms of rank conditions.

Let us introduce some notations used in the statement of the lemma below. Let Γ be the set of functions (u, c) in $\mathcal{H}_1$ and $\mathcal{F}_{\boldsymbol{\epsilon}}$ be the set of distributions $F_{\boldsymbol{\epsilon}}$ such that both sets satisfy Assumption 5. Consider another possible observationally equivalent function $\widetilde{\boldsymbol{m}}$, i.e., $\widetilde{m}^i(\boldsymbol{Q}, Z) \equiv N^{-1}s(X, Z) + N^{-1}X\widetilde{u}'(X, Z) - Q_i\widetilde{u}'(X, Z)$ for another function $\widetilde{u}$. For all $(\boldsymbol{q}, z)$, let $\frac{\partial \widetilde{\boldsymbol{m}}(\boldsymbol{q},z)}{\partial \boldsymbol{q}}$ and $\frac{\partial \boldsymbol{m}(\boldsymbol{q},z)}{\partial z}$ be the Jacobian matrices with respect to $\boldsymbol{q}$ and z, respectively, and $\partial \log(f_{\boldsymbol{\epsilon}}(\boldsymbol{m}(\boldsymbol{q}, z))/\partial \boldsymbol{\epsilon}$ be the $N \times 1$ gradient of $\log(f_{\boldsymbol{\epsilon}}(\boldsymbol{m}(\boldsymbol{q}, z)))$ with respect to $\boldsymbol{\epsilon}$, evaluated at $\boldsymbol{\epsilon} = \boldsymbol{m}(\boldsymbol{q}, z)$. Also, define

$$\Delta_{\boldsymbol{q}}(\boldsymbol{q}, z; \partial\boldsymbol{m}, \partial^2\boldsymbol{m}, \partial\widetilde{\boldsymbol{m}}, \partial^2\widetilde{\boldsymbol{m}}) \equiv \frac{\partial}{\partial \boldsymbol{q}}\log\left(\det\left(\frac{\partial \boldsymbol{m}(\boldsymbol{q}, z)}{\partial \boldsymbol{q}}\right)\right) - \frac{\partial}{\partial \boldsymbol{q}}\log\left(\det\left(\frac{\partial \widetilde{\boldsymbol{m}}(\boldsymbol{q}, z)}{\partial \boldsymbol{q}}\right)\right)$$

$$\Delta_z(\boldsymbol{q}, z; \partial \boldsymbol{m}, \partial^2 \boldsymbol{m}, \partial \widetilde{\boldsymbol{m}}, \partial^2 \widetilde{\boldsymbol{m}}) \equiv \frac{\partial}{\partial z} \log \left(\det \left(\frac{\partial \boldsymbol{m}(\boldsymbol{q}, z)}{\partial \boldsymbol{q}} \right) \right) - \frac{\partial}{\partial z} \log \left(\det \left(\frac{\partial \widetilde{\boldsymbol{m}}(\boldsymbol{q}, z)}{\partial \boldsymbol{q}} \right) \right),$$

Lemma 8. *Suppose that* $(\boldsymbol{m}, F_{\boldsymbol{\epsilon}}) \in \Gamma \times \mathcal{F}_{\boldsymbol{\epsilon}}$ *and* $\widetilde{\boldsymbol{m}} \in \Gamma$. *There exists* $\widetilde{F}_{\boldsymbol{\epsilon}} \in \mathcal{F}_{\boldsymbol{\epsilon}}$ *such that* $(\widetilde{\boldsymbol{m}}, \widetilde{F}_{\boldsymbol{\epsilon}})$ *is observationally equivalent to* $(\boldsymbol{m}, F_{\boldsymbol{\epsilon}})$ *if and only if for all* $\boldsymbol{q}, z$, *the rank of the matrix*

$$\begin{pmatrix} \left(\frac{\partial \widetilde{\boldsymbol{m}}(\boldsymbol{q},z)}{\partial \boldsymbol{q}}\right)' & \Delta_{\boldsymbol{q}}(\boldsymbol{q}, z; \partial \boldsymbol{m}, \partial^2 \boldsymbol{m}, \partial \widetilde{\boldsymbol{m}}, \partial^2 \widetilde{\boldsymbol{m}}) + \left(\frac{\partial \boldsymbol{m}(\boldsymbol{q},z)}{\partial \boldsymbol{q}}\right)' \frac{\partial log(f_{\boldsymbol{\epsilon}}(\boldsymbol{m}(\boldsymbol{q},z)))}{\partial \boldsymbol{\epsilon}} \\ \left(\frac{\partial \widetilde{\boldsymbol{m}}(\boldsymbol{q},z)}{\partial z}\right)' & \Delta_z(\boldsymbol{q}, z; \partial \boldsymbol{m}, \partial^2 \boldsymbol{m}, \partial \widetilde{\boldsymbol{m}}, \partial^2 \widetilde{\boldsymbol{m}}) + \left(\frac{\partial \boldsymbol{m}(\boldsymbol{q},z)}{\partial z}\right)' \frac{\partial log(f_{\boldsymbol{\epsilon}}(\boldsymbol{m}(\boldsymbol{q},z)))}{\partial \boldsymbol{\epsilon}} \end{pmatrix} \tag{4.3}$$

is N.

Next, we impose additional restrictions on the functional form of u in terms of additive separability of only characteristics z, and on the distribution of $\boldsymbol{\epsilon}$ as well as normalizations on u and c.

Assumption 6.

(i) $u(x, z) = u_1(x) + u_2(z)$.

(ii) For arbitrary $\boldsymbol{q} \in \mathcal{Q}^N$, *there exists* $z^0(\boldsymbol{q}) \in \mathcal{Z}$ *such that*

$$\frac{\partial f_{\boldsymbol{\epsilon}}(\boldsymbol{m}(\boldsymbol{q}, z^0(\boldsymbol{q})))}{\partial \epsilon_i} = 0 \ \forall i \in N \ \text{ and } \ \frac{\partial u_2(z)}{\partial z^0(\boldsymbol{q})} \neq 0,$$

(iii) $\underline{u}_1 \equiv u_1(\underline{x})$, $\underline{u}_1 = u_1'(\underline{x})$.

Condition (i) requires that there exists at least one element of Z satisfying the additive separability in u. It simply assumes that there exists at least one characteristic variable satisfying the additive separability, and it still allows for the general cases that other characteristics interact with the aggregate quantity in some unknown way because our identification arguments can be interpreted as given other characteristics, though based on the variation of z. It is worth noting that even though all characteristics are additively separable, it still captures the heterogenous marginal revenue because u is the marginal revenue function.

The first condition in (ii) impose restrictions on the distribution of $f_{\boldsymbol{\epsilon}}$, which is also used for the identification of a demand and supply model by [40], and the second condition is generally satisfied since it requires that the principal's marginal cost is not equal to agent's marginal effect of $u_2'(z)$ at $z = z^0(\boldsymbol{q})$. Condition (iii) is the normalization on the value of functions u at the lower bound $x = \underline{x}$.

Proposition 4. *Suppose* $(u, c) \in \Gamma$ *and* $F_{\boldsymbol{\epsilon}} \in \mathcal{F}_{\boldsymbol{\epsilon}}$. *Under Assumptions 5-6,* $u_1(\cdot)$ *is non-parametrically identified.*

The identification strategy takes several steps. First, based on the lemma 8 we identify the ratio u_1''/u_1' by combining conditions (i)-(ii) in Assumption 6, and hence identify $u(x, z)$ under additional normalization (iii).

4.3.2 Identification of $u_2(\cdot)$, $c(\cdot, z)$ and $F_{(\epsilon_1, \cdots, \epsilon_N)}(\cdot)$

The identification of $s(\cdot, z)$ is key to the identification of structural elements $u(\cdot, z)$ and $c(\cdot, z)$. Suppose that $s(\cdot, z)$ is identified. Then, $c(\cdot, z)$ can be identified if $u(\cdot, z)$ is identified. However, without more information, $s(\cdot, z)$ is identified up to a monotone transformation, as shown in [47]. In this view, [47] proposes some normalizations on $s(\cdot, z)$ to identify it. However, those normalizations are very restrictive in this model because $s(\cdot, z)$ is not the structural element, and the traditional linear specifications of $u(\cdot, z)$ and $c(\cdot, z)$ do not satisfy those normalizations. Indeed, those normalizations impose some functional relationship between $u(\cdot, z)$ and $c(\cdot, z)$. Since $u(\cdot, z)$ relies on the competition among agents while $c(\cdot, z)$ relies on the principal's productivity, it is not plausible to impose functional relationship between these two functions.

We will identify $s(\cdot, z)$ by using the identification results in [40, 42]. By exploring the (conditional) independence between Z and ξ and strictly monotonicity of $s(\cdot, z)$, we can identify the first partial derivative $s'(\cdot, z)$,[9] and then recover $s(\cdot, z)$ under some normalizations on $u(\underline{x}, z)$ and $c(\underline{x}, z)$.

Assumption 7.

(i) Z is an agent-specific variable such that $c(x, z) = c(x)$,

(ii) $\underline{u}_2 \equiv u_2(\underline{z})$, $\underline{c} \equiv c(\underline{x})$, $\underline{c}' \equiv c'(\underline{x})$, $\underline{c}'' \equiv c''(\underline{x})$.

Condition (i) is plausible because the cost function represents the productivity of the principal while Z affects the marginal revenue of the agent. Condition (ii) is normalizations on $u_2(\cdot)$ at $z = \underline{z}$, and $c(\cdot)$, $c'(\cdot)$ and $c''(\cdot)$ at $x = \underline{x}$.

Proposition 5. *Suppose $(u, c) \in \Gamma$ and $F_{\boldsymbol{\epsilon}} \in \mathcal{F}_{\boldsymbol{\epsilon}}$. Under Assumptions 5-7, $u_2(\cdot)$, $c(\cdot)$ and $F_{\boldsymbol{\epsilon}}(\cdot)$ are nonparametrically identified.*

4.3.3 Identification of $\gamma(z)$

Based on the identified functions u, c, and $F_{\boldsymbol{\epsilon}}$, we aim to identify the bargaining power function $\gamma(\cdot)$. According to the optimal condition (4.1), the optimal payment for each $i \in N$ is

$$
\begin{aligned}
P_i = {} & Q_i u(X, z) + \frac{\gamma(z)}{1 - \gamma(z) + N\gamma(z)} \left[c(X, z) - X u(X, z) - \sum_{j \neq i} Q_j \epsilon_j \right] \\
& + \frac{1 - 2\gamma(z) + N\gamma(z)}{1 - \gamma(z) + N\gamma(z)} Q_i \epsilon_i,
\end{aligned}
$$

Denote by $Y \equiv \sum_{i=1}^{N} P_i$ the aggregate optimal payment, we have

$$
Y = (1 - \gamma(z) + N\gamma(z))^{-1} \left[\gamma(z) N c(X, z) + (1 - \gamma(z)) \left(X c'(X, z) - u'(X, z) \sum_{i=1}^{N} Q_i^2 \right) \right], \tag{4.4}
$$

[9] See the formal proof in [40]. The identification is defined in terms of observational equivalence. Roughly, the main idea is to explore the restrictions that independence between X and Z imposes on $s(\cdot, z)$, and further, those restrictions can be expressed in terms of rank conditions.

In reality, the observed payment in the contract may deviate from the optimal payment int he sense that the observed payment $\widetilde{P}_i = P_i + \eta_i$ for each agent i, where η_i is a random term unknown to all players. One reason for the deviation is that η_i may represent the measurement error which is more plausible for the survey data. In addition, η_i could incorporate corruption, side payment, or political capture ([3]). Indeed, the deviation of observed payment from optimal payment is used not only in the nonparametric/semiparametric identification of contracts such as a contract model with adverse selection and moral hazard in [3], it is also adopted in the empirical analysis by using contract data. For example, [18] use the French transportation contracts to evaluate the renegotiation cost. Below we follow the same line in the identification of contract models to identify the bargaining power function $\gamma(z)$.

Assumption 8. *The observed payment $\widetilde{P}_i = P_i + \eta_i$ with $E(\eta_i|Z = z) = 0$ for each $i \in N$.*

The assumption is relatively weak in the sense that we do not impose any restrictions on the correlation between ϵ_i and η_i and hence allows for the correlation between ϵ_i and η_i. In particular, in some developing countries or emerging industries, there maybe some tacit collusion between the upstream firm and the downstream firms due to their institutional weakness, thus suggesting the correlation between between ϵ_i and η_i.

Under Assumption 8, $\widetilde{Y} \equiv \sum_{i=1}^{N} \widetilde{P}_i = Y + \zeta$ with $E(\zeta|Z = z) = 0$, where $\zeta \equiv$

$\sum_{i=1}^{N} \eta_i$. Hence, we can identify $\gamma(z)$ by $\gamma(z) = a_1(z)^{-1} a_2(z)$, where

$$\begin{aligned} a_1(z) &\equiv E\Big\{(N-1)Y - Nc(s^{-1}(\xi,z),z) + s^{-1}(\xi,z)c'(s^{-1}(\xi,z),z) \\ &\quad -[u'(s^{-1}(\xi,z),z)]^{-1}\sum_{i=1}^{N}\left[c'(s^{-1}(\xi,z),z) - u(s^{-1}(\xi,z),z) - \epsilon_i\right]^2 \Big| Z = z\Big\}, \\ a_2(z) &\equiv E\Big\{s^{-1}(\xi,z)c'(s^{-1}(\xi,z),z) \\ &\quad -[u'(s^{-1}(\xi,z),z)]^{-1}\sum_{i=1}^{N}\left[c'(s^{-1}(\xi,z),z) - u(s^{-1}(\xi,z),z) - \epsilon_i\right]^2 \Big| Z = z\Big\}. \end{aligned}$$

Now we summarize the identification result under passive beliefs.

Theorem 4. *Under the conditions in Lemma 6 and Assumptions 5-8, the model structure with passive beliefs* $\mathcal{S}_z \equiv [u(\cdot,z), c(\cdot,z), \gamma(z), F_{\boldsymbol{\epsilon}}(\cdot)]$ *is nonparametrically identified.*

The identification strategy applies to symmetric beliefs since the optimal conditions are similar for both beliefs.

Theorem 5. *Under the conditions in Lemma 7 and Assumptions 5-8, the model structure with symmetric beliefs* $\mathcal{S}_z \equiv [u(\cdot,z), c(\cdot), \gamma(z), F_{\boldsymbol{\epsilon}}(\cdot)]$ *is nonparametrically identified.*

4.4 Nonparametric Estimation

In this section, we propose the nonparametric estimators of the primitive functions by following the constructive nonparametric identification strategy, and then show the uniform consistency of these estimators under regular assumptions. Assume that the dataset $\{\boldsymbol{Q}_t, \boldsymbol{P}_t\}_{t=1}^{T}$ is generated by the same equilibrium, where $\boldsymbol{Q}_t = (Q_{t,1}, \cdots, Q_{t,N})$, $\boldsymbol{P}_t = (P_{t,1}, \cdots, P_{t,N})$, and T is the sample size. As in the identification, the estimation procedures under both belief structures are similar and without loss of generality, we focus on the estimation under passive beliefs.

4.4.1 Nonparametric Estimators

The nonparametric estimation takes two steps. First, we adopt the sieve method to estimate $c(\cdot,\cdot)$, $u'(\cdot,\cdot)$ and $\gamma(\cdot)$ based on the equation of optimal payment Y. Second, we estimate the pseudo random shock vector $\boldsymbol{\epsilon}$ and thereby estimate the joint distribution function $F_{\boldsymbol{\epsilon}}$.

Recall that $\widetilde{Y} = Y + \zeta$ implied by Assumption 8, the identification of bargaining power function is obtained under the weak exogeneous condition $E(\zeta|Z) = 0$ without imposing stronger condition $E(\zeta|\boldsymbol{Q}, Z) = 0$. To my knowledge, under $E(\zeta|\boldsymbol{Q}, Z) \neq 0$ there is no existing literature on the consistent estimators of $(u, c, \gamma, F_{\boldsymbol{\epsilon}})$. Hence, we adopt the sieve method to solve the potential endogeneity of $\boldsymbol{Q}$ by assuming that there exists an instrumental vector $Z_q \in \mathcal{Z}_q \subset \mathcal{R}^N$ such that $E(\zeta|V) = 0$, where $V \equiv (Z', Z_q')' \in \mathcal{Z} \times \mathcal{Z}_q$, and $\mathcal{Z} \times \mathcal{Z}_q$ is a compact subset of $\mathcal{R}^{d_z+N}$. In our used data on stents, one plausible instrument vector is the quantity of the balloon-tipped catheter of each hospital. During the process of an angioplasty the doctor threads a balloon-tipped catheter from a peripheral access point to the heart, and hence it is reasonable to assume that the quantity of catheters is to a large extent correlated to the quantity of stents while it is uncorrelated to the payment error of stents. [10]

Following [41], we consider a typical space of smooth functions, the Hölder space, to which the unknown true functions belong. For any generic vector $w \in \mathcal{W} \subset \mathbf{R}^{d_w}$ with $d_w > 0$, let $a = (a_1, \cdots, a_{d_w})^T$ and denote the $(a_1 + \cdots + a_{d_w})$th derivative of $g(w)$ by

$$\nabla^a g(w) \equiv \frac{\partial^{a_1+a_2+\cdots+a_{d_w}} g(w)}{\partial w_1^{a_1} \cdots \partial w_{d_w}^{a_{d_w}}}.$$

[10] A balloon catheter is a type of "soft" catheter with an inflatable "balloon" at its tip which is used during a catheterization procedure to enlarge a narrow opening or passage within the body. The deflated balloon catheter is positioned, then inflated to perform the necessary procedure, and deflated again in order to be removed. Some common uses include: angioplasty or balloon septostomy, via cardiac catheterization (heart cath); tuboplasty via uterine catheterization; pyeloplasty using a detachable inflatable balloon stent positioned via a cystocopic transvesicular approach.

The Hölder space $\Lambda^{\iota}(\mathcal{W})$ of order ι is a space of functions $g : \mathcal{W} \mapsto \mathbf{R}$ such that the first $\underline{\iota}$ derivative is bounded, where $\underline{\iota}$ denote the largest integer satisfying $\iota > \underline{\iota}$, and the $\underline{\iota}$ derivative are Hölder continuous with the exponent $\iota - \underline{\iota} \in (0, 1]$, i.e.,

$$\max_{a_1+a_2+\cdots+a_{d_w}=\underline{\iota}} |\nabla^a g(w) - \nabla^a g(w')| \leq \text{const.}(||w - w'||_E)^{\iota-\underline{\iota}}$$

for all w, $w' \in \mathcal{W}$, where $||\cdot||_E$ denotes the Euclidean norm. Furthermore, the Hölder space becomes a Banach space with the following Hölder norm

$$||g||_{\Lambda^{\iota}} = \sup_{w\in\mathcal{W}} |g(w)| + \max_{a_1+a_2+\cdots+a_{d_w}=\underline{\iota}} \sup_{w\neq w'} \frac{\nabla^a g(w) - \nabla^a g(w')}{(||w - w'||_E)^{\iota-\underline{\iota}}}.$$

The Hölder space denoted by $\Lambda^{\iota}(\mathcal{W})$ incurs a Hölder ball $\Lambda_b^{\iota}(\mathcal{W}) \equiv \{g \in \Lambda_b^{\iota}(\mathcal{W}) : ||g||_{\Lambda^{\iota}} \leq b < \infty\}$. It is known that power series, Fourier series, splines and wavelets all can approximate functions in $\Lambda_b^{\iota}(\mathcal{W})$ well. Let $W = (X, Z')'$ and $\mathcal{W} = \mathcal{X} \times \mathcal{Z}$. For some integer $\iota > (1 + d_z)/2$, we assume the unknown true functions $\varpi_0 = (u_0', c_0, \gamma_0)$ belong to the set of functions $\mathcal{F} \equiv \mathcal{C} \times \mathcal{U} \times \Upsilon$, where

$$\begin{aligned}
\mathcal{C} &\equiv \{c(\cdot) : c(\cdot) \in \Lambda_b^{\iota}(\mathcal{X}) \text{ and Assumption 6 holds}\} \\
\mathcal{U} &\equiv \{u_1'(\cdot) : u_1'(\cdot) \in \Lambda_b^{\iota}(\mathcal{X}) \text{ and Assumption 6 holds}\} \\
\Upsilon &\equiv \{\gamma(\cdot) : \gamma(\cdot) \in \Lambda_b^{\iota}(\mathcal{Z})\}.
\end{aligned}$$

Then, we use the sieve method to estimate the unknown primitive functions with a univariate B-spline of order $r > \iota$ given by

$$B_r(u) = \frac{1}{(r-1)!} \sum_{i=0}^{r} (-1)^i C_r^i [\max(0, u - i)]^{r-1},$$

to construct a spline-wavelet sieve basis $\{2^{k/2} B_r(2^k W_l - i) : i = 0, \pm 1, \pm 2, \cdots, k =$

$0, \cdots, K_T\}$ for each $l = 1, \cdots, d_w$ to approximate functions in $\Lambda^{\iota}(\mathcal{X} \times \mathcal{Z})$, where K_T is the smoothing parameter depending on the sample size T. [11] In the sieve approximations, we will replace the space $\mathcal{F}$ with the finite-dimensional space $\mathcal{F}_T \equiv \mathcal{C}_T \times \mathcal{U}_T \times \Upsilon_T$, where

$$\begin{aligned}
\mathcal{C}_T &\equiv \{c(x,z) = p^{K_T}(x)'\delta_c \text{ for all } \delta_c \text{ such that } c \in \mathcal{C}\} \\
\mathcal{U}_T &\equiv \{u_1'(x) = p^{K_T}(x)'\delta_u \text{ for all } \delta_u \text{ such that } u_1' \in \mathcal{U}\} \\
\Upsilon_T &\equiv \{\gamma(z) = p^{K_T}(z)'\delta_\gamma \text{ for all } \delta_\gamma \text{ such that } \gamma \in \Upsilon\}.
\end{aligned}$$

By assuming the homoscedasticity of $\eta_{t,i}$ ($t = 1 \cdots, T, i = 1, \cdots, N$), the sieve minimum distance (SMD) estimators of $(\hat{c}(\cdot), \hat{u}'(\cdot), \hat{\gamma}(\cdot))$ are defined as

$$(\hat{c}(\cdot), \hat{u}_1'(\cdot), \hat{\gamma}(\cdot)) = \underset{\varpi \equiv (c(\cdot), u_1'(\cdot), \gamma(\cdot)) \in \mathcal{F}_T}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T} \widehat{m}(V_t, \varpi)'\widehat{m}(V_t, \varpi), \tag{4.5}$$

where $\widehat{m}(v, \varpi)$ is the sieve estimator of $m(v, \varpi) \equiv \int \rho(\tilde{y}, \boldsymbol{q}, z; \varpi) dF_{\widetilde{Y}, \boldsymbol{Q}|V=v}(\tilde{y}, \boldsymbol{q})$,

$$\begin{aligned}
\rho(\widetilde{Y}, \boldsymbol{Q}, Z, \varpi) &\equiv \widetilde{Y} - \frac{N\gamma(Z)}{1-\gamma(Z)+N\gamma(Z)}c(X) - Xc'(X)\frac{1-\gamma(Z)}{1-\gamma(Z)+N\gamma(Z)} \\
&+ \frac{(1-\gamma(Z))}{1-\gamma(Z)+N\gamma(Z)}u_1'(X)\sum_{j=1}^{N} Q_j^2,
\end{aligned}$$

and

$$\widehat{m}(V, \varpi) = \sum_{t=1}^{T} \rho(\widetilde{Y}_t, \boldsymbol{Q}_t, Z_t, \varpi) p^{K_T}(V_t)'(P'P)^{-1} p^{K_T}(V_t),$$

where $P \equiv (p^{K_T}(V_1), \cdots, p^{K_T}(V_T))'$. Indeed, the sieve estimator $\widehat{m}(V, \varpi)$ is simply the ordinary least square estimation by regressing $\rho(\widetilde{Y}_t, \boldsymbol{Q}_t, Z_t, \varpi)$ on $p^{K_T}(V_t)'$ and hence it can be interpreted as a GMM or nonlinear 2SLS estimator (see [41]).

[11] Suppose the smoothing parameter is K_{xT} for $c(x)$ and $u_1(x)$ and is K_{zT} for $\gamma(z)$. Since d_z can be greater than 1, we employ a tensor-product linear sieve basis that is simply the product of univariate sieves such that the constructed tensor-product sieve basis for $\gamma(z)$ is $K_{1zT} \times 1$ vector $p^{K_{1zT}}(Z)$ with $K_{1zT} = (2^{K_{zT}})^{d_z}$. For simplicity, we assume that $K_{1zT} = K_{xT} = K_T$, though this can be easily relaxed.

In the second step, we propose the nonparametric estimators of $u(x,z)$ and $F_{\boldsymbol{\epsilon}}(\cdot)$. Since $\xi = Nc'(X,Z) - Nu_1(X) - Nu_2(Z) - Xu_1'(X)$, under the assumption that $\mathbf{E}[\epsilon_i|Z = z] = 0$, we obtain the kernel estimator of $u_2(z)$. The conditional zero mean assumption is plausible when $u(\cdot)$ is interpreted as the average marginal revenue in the market. Let $K : \mathcal{R}^{d_z} \to \mathcal{R}$ be a kernel function and h_T is the bandwidth depending on the sample size T. Then

$$\widehat{u}_2(z) = \left[\sum_{t=1}^{T} [\widehat{c}'(X_t) - \widehat{u}_1(X_t) - N^{-1}X_t\widehat{u}_1'(X_t)] K\left(\frac{Z_t - z}{h_T}\right)\right]\left[\sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right)\right]^{-1}.$$

Therefore, we can construct a sample of pseudo random vector through $\widehat{\epsilon}_{t,i} = \widehat{c}_1'(X_t) - \widehat{u}_1(X_t) - \widehat{u}_2(Z_t) - \widehat{u}_1'(X_t)Q_{t,i}$, for each $t = 1, \cdots, T$ and $i = 1, \cdots, N$.

Finally, we propose the kernel estimator of joint distribution $f_{\boldsymbol{\epsilon}}$ by using the sample of pseudo random vectors. Let $\boldsymbol{e} = (e_1, \cdots, e_N) \in \mathcal{S}_{\boldsymbol{\epsilon}}$, where $S_{\boldsymbol{\epsilon}}$ is a bounded set excluding the boundary of $\mathcal{E}^N$, $\widehat{f_{\boldsymbol{\epsilon}}}(\boldsymbol{e}) = (Th_\epsilon^N)^{-1}\sum_{t=1}^{T}\prod_{i=1}^{N} k_\epsilon\left(\frac{e_i - \widehat{\epsilon}_{t,i}}{h_\epsilon}\right)$, where $k_\epsilon(\cdot)$ is a PDF, and h_ϵ is the bandwidth depending on the sample size T.

4.4.2 Uniform Consistency

In this subsection we first establish the uniform consistency of $\widehat{c}(x,z)$, $\widehat{u}'(x,z)$, and $\widehat{\gamma}(z)$, under the following sufficient assumptions.

Assumption 9. *The data $\{(\boldsymbol{Q}_t, \boldsymbol{P}_t, Z_t)\}_{t=1}^{T}$ are $i.i.d.$.*

This $i.i.d.$ condition can be easily relaxed for the results on consistency and rate of convergence ([41]).

Assumption 10. *(i) V is a compact connected subset of $\mathcal{R}^{N+d_z}$ with Lipschitz continuous boundary, and f_V is bounded and bounded away from zero. (ii) $\underset{1\leq k\leq K_T}{max} E[|p_k(V)|^2] \leq$ const., and the smallest eigenvalue of $E[p^{K_T}(\cdot) \times p^{K_T}(\cdot)']$ is bounded away from zeros for all K_T. (iii) There is a sequence of measurable functions $\{\overline{\rho}_T(Y, \boldsymbol{Q}, Z)\}_{T=1}^{\infty}$ and a finite* *constant $M_1 > 0$ such that $|\rho(\widetilde{Y}, \boldsymbol{Q}, Z; \Pi_T \varpi_0)| \leq \overline{\rho}_T(\widetilde{Y}, \boldsymbol{Q}, Z)$ and $E[\overline{\rho}_T(\widetilde{Y}, \boldsymbol{Q}, Z)^2|V] \leq M_1$, and there is $p^{K_T}(V)'\pi$ such that $E\{[m(V, \Pi_T\varpi_0) - p^{K_T}(V)'\pi]^2\} = O(K_T^{-2\iota/(N+d_z)})$. (iv) $K_T = o(T^{1/2})$.*

Assumption 10-(iv) requires a relatively large sample size or a relatively small dimension of Z.

Assumption 11. *$E[m(V, \varpi)'m(V, \varpi)]$ is continuous at ϖ_0 under the sup norm.*

Assumption 12. *(i) $K_T < \infty$ and $K_T \to \infty$ as $T \to \infty$. (ii) Assumptions 6 hold for ϖ in a neighborhood of ϖ_0 in the sup norm $||\cdot||_s$.*

Assumption 12-(ii) guarantees that $\mathcal{F}$ is compact under the sup norm.

Lemma 9. *Let $(\hat{c}(\cdot), \hat{u}'(\cdot), \hat{\gamma}(\cdot))$ defined in (4.5) be the SMD estimator. Suppose $E[\eta_{t,i}^2|V] = \sigma_\eta^2 I$ with $\sigma_\eta > 0$, Assumptions 9-12 hold, then $\sup_{(x,z)} |\hat{c}(x) - c_0(x)| = o_p(1)$, $\sup_{(x,z)} |\hat{u}_1'(x) - u_{1,0}'(x)| = o_p(1)$, and $\sup_{z} |\hat{\gamma}(z) - \gamma_0(z)| = o_p(1)$.*

Given the uniform consistency result, we restrict our attention to a shrinking $||\cdot||_s$ neighborhood around ϖ_0. Let

$$\mathcal{F}_s \equiv \{\varpi \in \mathcal{F} : ||\varpi - \varpi_0||_s \leq \mu, ||\varpi||_s \leq M_2, \} \text{ and } \mathcal{F}_{sT} \equiv \mathcal{F}_s \cap \mathcal{F}_T$$

for some positive finite constants M_2, and a sufficiently small positive μ such that $Pr(\widehat{\varpi} \notin \mathcal{F}_s) < \mu$. For the purpose of establishing a rate of convergence under the sup norm, we treat $\mathcal{F}_s$ as the new space and $\mathcal{F}_{sT}$ as its sieve space. We introduce the pseudometric on $\mathcal{F}_s$ that could be weaker than $||\cdot||_s$. Let the first pathwise derivative in the direction $[\varpi - \varpi_0]$ evaluated at ϖ_0 as

$$\frac{dm(V, \varpi)}{d\varpi}[\varpi - \varpi_0] \equiv \frac{dE[\rho(Y, \boldsymbol{Q}, Z, (1-\tau)\varpi_0 + \tau\varpi)|V]}{d\tau}\Big|_{\tau=0},$$

and

$$||\varpi_1 - \varpi_2||^2 \equiv E\left[\left(\frac{dm(V,\varpi_0)}{d\varpi}[\varpi_1 - \varpi_2]\right)'\left(\frac{dm(V,\varpi_0)}{d\varpi}[\varpi_1 - \varpi_2]\right)\right]$$

Assumption 13. *(i) $\mathcal{F}_s$ and $\mathcal{F}_{sT}$ are convex, and there is a finite constant $M_3 > 0$ such that $||\varpi - \varpi_0|| \leq M_3||\varpi - \varpi_0||_s$ for all $\varpi \in \mathcal{F}_{os}$. (ii) There are finite constants $M_4, M_5 > 0$ such that $||\varpi - \varpi_0||^2 \leq M_4 E[m(V,\varpi)'m(V,\varpi)]$ holds for all $\varpi \in \mathcal{F}_{osT}$; and $M_5 E[m(V,\Pi_T\varpi_0)'m(V,\Pi_T\varpi_0)] \leq ||\Pi_T\varpi_0 - \varpi_0||^2$.*

Assumption 13-(i) implies that the pseudometric $||\varpi - \varpi_0||$ is weaker than $||\varpi - \varpi_0||_s$. (ii) implies that the weaker pseudometric $||\varpi - \varpi_0||$ is Lipschitz continuous with respect to the population criterion function $E[m(V,\varpi)'m(V,\varpi)]$ for all $\varpi \in \mathcal{F}_{sT}$.

Proposition 6. *Let $(\hat{c}(\cdot), \hat{u}_1'(\cdot), \hat{\gamma}(\cdot))$ defined in (4.5) be the SMD estimator. Under the same assumptions in Lemma 9 and Assumption 13, then $\sup_{(x,z)} |\hat{c}(x,z) - c(x,z)| = O_p(a_T)$, $\sup_{(x,z)} |\hat{u}_1'(x) - u_{1,0}(x)| = O_p(a_T)$, and $\sup_{z} |\hat{\gamma}(z) - \gamma(z)| = O_p(a_T)$, where $a_T = a_{1T} + a_{2T}$, $a_{1T} = (K_T)^{-\iota/(1+d_z)}$, $a_{2T} = \sup_{\varpi \in \mathcal{F}_{osT}: ||\varpi - \Pi_T\varpi_0|| \leq max\{\delta_{m,T}, ||\varpi_0 - \Pi_T\varpi_0||\}} ||\varpi - \Pi_T\varpi_0||_s$ with $\delta_{m,T}^2 = max\left(T^{-1}K_T, K_T^{-2\iota/(N+d_z)}\right)$*

The proof follows Theorem 4.1 in [39] and omitted.

Second, we prove the uniform consistency of $\widehat{u}_2(z)$ and $\widehat{\epsilon}_{t,i}$ under standard smoothing conditions of the density of z.

Assumption 14. *(i) The density f_Z is differentiable and $u_2(z)$ is twice differentiable, and that the derivative functions all satisfy the Lipschitz condition $|a(z) - m(z')| \leq C_1|z - z'|$ for some $C_1 > 0$, where $a(z) = u_2''(z)$ or $f_Z'(z)$. (ii) $\inf_{z \in \mathcal{Z}_{int}} f_Z(z) \geq \delta > 0$, where $\mathcal{Z}_{int}$ exclues the boundary range of $\mathcal{Z}$. (iii) The kernel $K(\cdot)$ is symmetric, bounded, and has compact support. Define $H_l(z) = |z|^l K(z)$, we assume that $|H_l(z) - H_l(z')| \leq C_2|z - z'|$ for some $C_2 > 0$ for all $0 \leq l \leq 3$. (iv) As $T \to \infty$, $h_T \to 0$ and $Th_T(lnT)^{-1} \to \infty$.*

Assumption 14 are standard assumptions on uniform convergence of kernel estimators.

Proposition 7. *Suppose that Assumptions in proposition 6 and Assumption 14 hold. Then,* $\sup_{z\in\mathcal{Z}_{int}} |\widehat{u}_2(z) - u(z)| = O_p(b_T) = o_p(1)$, *where* $b_T = ln(T)^{1/2}(Th_T)^{-1/2} + h_T^2$.

Let

$$\widetilde{u}_2(z) = \left[\sum_{t=1}^{T} \left[c'(X_t) - u_1(X_t) - N^{-1}X_t u_1'(X_t)\right] K\left(\frac{Z_t - z}{h_T}\right)\right] \left[\sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right)\right]^{-1}.$$

Then

$$\sup_{z\in\mathcal{Z}_{int}} |\widehat{u}_2(z) - u(z)| \leq \sup_{z\in\mathcal{Z}_{int}} |\widehat{u}_2(z) - \widetilde{u}(z)| + \sup_{z\in\mathcal{Z}_{int}} |\widetilde{u}_2(z) - u(z)| = O_p(a_T + b_T) = o_p(1).$$

where the first rate a_T follows proposition 6, and the second rate b_T follows the standard result on kernel estimators An immediate consistency result of Proposition 7 is $\sup_{t,i} |\widehat{\epsilon}_{t,i} - \epsilon_{t,i}| = O_p[\max(a_T, b_T)] = o_p(1)$ because $\widehat{\epsilon}_{t,i} = \widehat{c}'(X_t) - \widehat{u}(X_t, Z_t) - \widehat{u}_1'(X_t)Q_{t,i}$, for any t and i.

Assumption 15. *(i) The bounded kernel* $k_\epsilon \in \mathcal{L}_1$ *satisfies* $|k_\epsilon(e) - k_\epsilon(\widetilde{e})| \leq C_\epsilon |e - \widetilde{e}|$ *for all* $e, \widetilde{e}$ *on* $\mathcal{E}^N$ *with some constant* $C_\epsilon > 0$*; (ii) The bandwidth* $h_\epsilon \to 0$ *such that* $Th_\epsilon^N(lnT)^{-1} \to \infty$ *and* $(a_T + b_T)T(lnT)^{-1} \to 0$ *as* $T \to \infty$

Proposition 8. *Under Assumptions in proposition 7 and Assumption 15, we have*

$$\sup_{\boldsymbol{e}\in\mathcal{S}_\epsilon} \left|\widehat{f}_{\boldsymbol{\epsilon}}(\boldsymbol{e}) - f_{\boldsymbol{\epsilon}}(\boldsymbol{e})\right| = O_p\left\{(a_T + b_T)h_\epsilon^{-N} + \left[(Th_\epsilon^N)^{-1} lnT\right]^{1/2} + h_\epsilon^2\right\} = o_p(1).$$

4.5 Empirical Application

We now apply our method to the coronary stents contracts in the United States. As in the theoretical model, the manufacturer of stents negotiate with each hospital, specifying the payment for a given quantity of stents in each bilateral contract. Due to the secrecy of contracts and the competition among hospitals, this unique dataset is particularly suitable for our model. Our empirical findings show that passive beliefs fit the dataset better than symmetric beliefs. More relevantly, the counterfactual result shows that if the agent's bargaining power increases, under passive beliefs the stent's price decreases by some reasonable amount, while under symmetric beliefs the decrease of price is unreasonable.

4.5.1 Data and Background on Coronary Stent

The coronary stent is a critical medical device used in angioplasty to treat the coronary artery disease (i.e., blockages in the arteries surrounding the heart) which might cause heart attack or even death. The traditional "bare metal stent" (BMS) has the drawback that scar tissue growth can lead to significant renarrowing of the artery. For this reason, the "Drug-eluting stent" (DES) was introduced in the early 2000s as an improvement over BMS by coating the stent with a drug that discourages scar tissue growth and significantly reduces the incidence of renarrowing. The global sales of coronary stents was \$8.1 billion in 2014 and is expected to grow to roughly \$10.6 billion in 2019, and the North American market reached \$3.1 billion in 2014 and is expected to grow to \$4.4 billion in 2019.[12]

The dataset is from Millennium Research Group's (MRG) *MarketTrack* survey of hospitals in US from 2006 to 2014, and its goal is to provide representative estimates of market shares and prices by US region (Northeast, Midwest, South, West). The dataset consists of quantities and prices as well as hospital's characteristics (census region, teaching/nonteaching, public/private) at the product-hospital-month level.[13] The stent market

[12] Source: BCC Research 2015, "Stents: Technologies and Global Markets."

[13] The four regions are based on the regional divisions used by the US Census Bureau. A teaching hospital is a facility that is associated with a medical school and provides clinical training for health care professionals. A public hospital is a facility that is funded by the state or federal government, while a private hospital is funded by patients and insurers. Note that a private hospital can be for-profit or not-for-profit. Non-profit hospitals are mostly funded by charity, religion or research/educational funds. The hospital industry in the United States includes a mix of ownership forms. Non-profit hospitals are the most common type, but for-profit and government hospitals also play substantial roles. Nonprofit hospitals do not pay federal income or state and local property taxes, and in return they benefit the community. For-profit hospitals, or alternatively investor-owned hospitals, attempt to garner a profit for their shareholders. Because public hospitals are pub-

is dominated by four firms: the Abbott Vascular (formerly Guidant) division of Abbott Laboratories, Boston Scientific, Johnson and Johnson's Cordis division, and Medtronic.[14]

The monthly sample can be representative of the contracting game despite the monthly frequency. The sample pool consists of 160 hospital labs. Because the identities of hospitals vary in each month's observation in the dataset, the sample provides substantial variations of contracts, suggesting the representativeness of the sample. Indeed, the goal of the dataset by MRG is to produce representative estimates of the distribution of market shares and prices in the stent market.

The empirical analysis will focus on the subsample of the northeastern region in terms of the intensity of competition among hospitals in these four regions. On the one side, the area of northeastern region is much smaller than other three regions and patients are very sensitive to travel time ([60]; [73]). On the other side, Figure 4.1 shows that the market share of quantities in northeastern region is comparable to other three regions on a monthly basis. Hence, the most competitive northeast region is most relevant to our model. Actually, the dataset collected by MRG is to produce representative estimates of the distribution of market shares and prices by region, which is the relevant unit of observation for this study.

4.5.2 Descriptive Analysis

Figure 4.2 provides the basic patterns of two key variables quantities and prices. As the left panel indicates, during 2006 the DES market size decreased dramatically while the BMS market size increased moderately. This is caused by a study concerning the safety of DES, though one year later it became clearer that DES were not as dangerous as the study suggested. Since 2007, Both DES and BMS total quantities are relatively stable over time,

licly funded and not for profit, they are usually a lot more affordable than private hospitals (see more details at http://www.npinstitute.com/public-vs-private-hospitals-s/1852.htm).

[14]Because the dataset used in this paper is sold as market research to the device manufacturers, hospitals are anonymous, which might prevent other analysis involving the identities of hospitals.

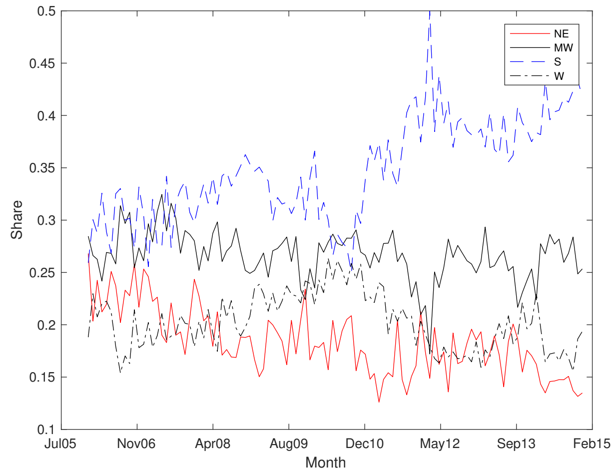

Figure 4.1: Market shares of quantities for each region over time

though the BMS market size is slightly decreasing from 2013.

Based on the right panel, average prices of BMS and DES have a markedly decreasing trend, and in particular, DES price decreased by about 30 percent over time, which maybe explained by a more competitive stent market over time. During the sample period more manufacturers entered the stent market, resulting in decreasing prices.[15] In addition, the price of DES is much higher than BMS due to the advantages of DES over BMS. Therefore, we will use type-month detrended prices in the empirical analysis to control for both the effect of stent types on prices and the effect of competition among manufacturers on prices, where detrended prices are obtained from a regression of prices on dummy for types and time trend. Accordingly, detrended prices can match the model well in the sense that the four manufacturers can be considered to behave as a single manufacturer.

We will show the relevance of the dataset to our model in terms of competition among hospitals and negotiated prices, respectively. First, hospitals compete for patients with

[15] See "Positive Studies Boost Stent Manufacturers as Market Competition Heats Up" at http://www.mddionline.com/article/positive-studies-boost-stent-manufacturers-market-competition-heats.

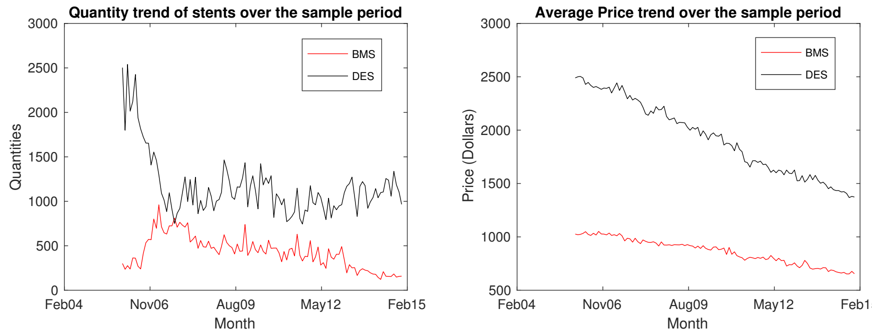

Figure 4.2: Total quantities and average price trend over time

coronary artery diseases by increasing the quality of service. Using data in Pennsylvania during the years 1995-2004, [73] measure the degree of competition among hospitals for patients with coronary artery diseases in terms of predicted market shares, and find that hospitals in more competitive markets achieved lower mortality among severely ill patients by increasing the quality of service.

Second, the variation of type-month detrended prices of coronary stents is still substantial across hospitals due to the following reasons. As Figure 4.3 displays, there is a large variation of detrended prices across hospitals, with a standard deviation of $\$125$ and mean of $\$1,212$, for a coefficient of variation of 0.10. Further, the regression of detrended prices on dummies for hospitals indicates that hospital fixed effects explain about one half of price variation, with an adjusted $R^2 = 0.488$. One explanation is that since the payment in the contract is a lump-sum for all quantities of stents and the payment is nonlinear in the quantity and thus the average price varies with the quantity ([74]). Moreover, the quantity varies with the unobserved marginal revenue term ϵ and there seems large variation of marginal revenue across hospitals.[16] Hence, the large variation of detrended prices maybe due in part to the large variation of hospitals' marginal revenues.

[16]There is a considerably large variation of margin for the cardiac valve replacement surgery across hospitals in U.S. ([75]). Also see the large variation of operating margin across hospitals in New Hampshire at https://www.nh.gov/insurance/lah/documents/umms.pdf

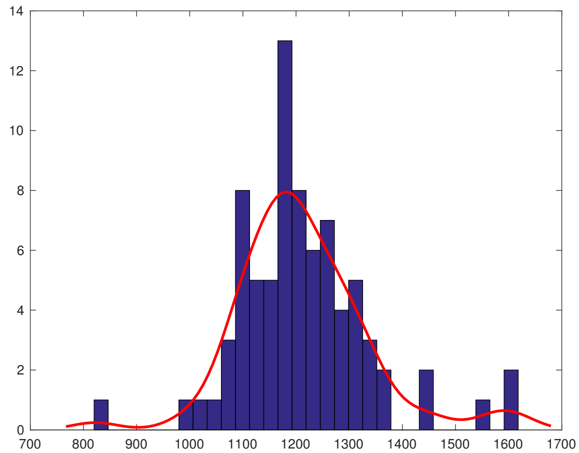

Figure 4.3: Distribution of detrended prices across hospitals

Another explanation is from the heterogeneous bargaining powers across hospitals. Since the characteristics of hospitals is different in the dataset, in which the bargaining power may depend on those characteristic, heterogeneous bargaining power could also contribute to the substantial variation of detrended prices. In reality, the price of stents is determined through private negotiation between a hospital administrator and a representative of the manufacturer, and different hospitals negotiate different prices for the same stent at the same point of time.[17] It is worth noting that even though hospitals typically have limited information regarding other hospitals' past contracts,[18] they can not observe the current contracts of other hospitals. In addition, those limited information would reduces the un-

[17]Hospitals typically rely on the service of group purchasing organizations (GPOs) to negotiate contracts for many products. In terms of the degree of involvement of GPO, there are two dominant models for hospital engagement with GPOs. The high involvement of GPO implies that hospitals primarily use the GPO prices to purchase medical devices while the low involvement of GPO implies that hospitals use GPO prices as a starting point for direct hospital-manufacturer negotiations. For physician preference items hospitals usually use GPO prices as a starting point for direct hospital-manufacturer negotiations, where physician preference items are those products whose demand is determined in large part by preferences of brand-loyal physicians. Accordingly, [76] adopts the direct hospital-manufacturer negotiation in the coronary stent market.

[18]Policy-makers have concerned the secrecy of contracts in medical device industry. In 2014, Senator Angus King of Maine added an amendment to a tax bill that would increase price transparency for medical devices.

certainty in the manufacturer's cost or bargaining parameter, thus implying the realistic assumption of common knowledge.[19]

The specification of payoff of agent in the model fits the dataset well. First, because the major source of hospital's revenue comes from the reimbursement rate from patient's insurer to hospital, and these rates depends on the aggregate demand of hospitals in the United States ([59]; [60]), the specification of $u(x)$ makes sense. Second, ϵ_i represents the unobserved marginal revenue, which is affected by various factors, including the heterogenous reimbursement policies ([79]; [80]; [78]; [81]),[20] the unobserved quality of hospital i's service ([60]), the unobserved hospital i's cost-related terms ([81]), and the patient characteristics ([82]). Third, it is plausible that $E[\eta_i|\boldsymbol{Q}, Z] = 0$ needed for the identification of parametric specifications for the following reasons. In our survey dataset η_i could be interpreted as the measurement error; and the reimbursement rate of the stent-related surgeries does not depend on the use of stents in the coronary stent industry ([76]), thus suggesting that the risk of potential collusion between the manufacturer and hospitals might be ignorable in the coronary stent market.

4.5.3 Empirical Strategy

The two primary goals of our application are (1) to test for the equivalence of passive beliefs and symmetric beliefs by using the estimates of the parameters in the cost function, marginal revenue function, and bargaining power function under two belief system, respectively; (2) to assess the cut-down of prices that would be achieved if hospitals'

[19]Based on the claim by [77] that the HMO knows as much (and often more) about the target hospital's cost structure as the hospital's management, [59] assume that both the hospitals and the HMO know the profitability of different networks and the cost structures of all hospitals (see similar assumptions in [78]). Similarly, in the manufacturer-hospital negotiation [76] assumes that both the hospitals and the manufacturer know the profitability of different hospitals and the manufacturer's cost function. This paper follows this line in the health care market, i.e., the common knowledge of $\boldsymbol{\epsilon}$ among hospitals and manufacturers.

[20]Since the reimbursement negotiated rates between the hospital and its insurer are determined by the heterogenous bargaining power of hospitals ([79]; [80]), ϵ_i is affected by hospital i's unobserved idiosyncratic terms related to the hospital's bargaining power ([81]). Usually the reimbursements from private insurers are generally negotiated as a markup on Medicare rates across all procedures performed at the hospital.

bargaining powers increase by the same amount through relevant policies, as well as the heterogeneous effect of increasing bargaining power on the decrease of prices. Recall that the identification argument in Section 3 is presented explicitly conditional on given characteristics $Z = z$ and by suppressing the vector W_1. Since in our data W_1 could be the dummy vector for four regions in US, to be consistent with the identification strategy, in this section we allows for heterogeneity $Z = z$ among hospitals conditional on the northwest region. Acknowledging the limited sample size, we adopt a parametric specification of model primitives to make estimation feasible.

It is worth noting that in this data, the unobservable ϵ may capture monthly variation in demand occurs due to changes in doctor preferences (as new studies are released and device salespeople spread the word), changes in unobserved patient characteristics ([80]; [76]), or the component of marginal cost of hospitals([78]). In addition, the assumption that disagreements of hospitals and manufacturer are zero is plausible in two ways. Due to brand-loyalty of physicians in the stent market ([76]), profits of non-stent treatment (typically a suggested diet and exercise regimen) can be normalized to be zero. The market breakdown assumption can be a good approximation based on the fact of DES scare in 2006 which is caused by a study concerning the safety of DES resulted in less DES usage and less stenting overall, though one year later it became clearer that DES were not as dangerous as the study suggested.[21]

Recall that in Section 4 we presented a nonparametric approach to estimate the primitive functions (u, c, γ). A practical issue of such a nonparametric approach in our application is that a large sample size is required due to the slow rate of convergence of nonparametric estimators. Therefore, we take an alternative approach by parameterizing the model primitives.[22]

[21] See "Embers still smoldering from the 2006 ESC firestorm, as experts mull DES safety and efficacy" at http://www.medscape.com/viewarticle/708137.

[22] In the coronary stent industry one plausible variable used in the nonparametric identification is the pop-

We specify hospital i's revenue function by taking account into their characteristics

$$\pi^h \equiv Q_i(\alpha_0 + \alpha_1 X + \alpha_2 X^2 + Z_i'\alpha_3 + Z_i'\alpha_4 X + \epsilon_i) - P_i,$$

where Q_i and P_i are, respectively, the quantity and payment in hospital i's contract for each $i \in N$, $X \equiv \sum_{i=1}^N Q_i$ is the aggregate quantity, Z_i is the vector of hospital i's characteristics. The manufacturer's cost function is specified as

$$\pi^m \equiv \sum_{i=1}^N P_i - (\beta_0 + \beta_1 X + \beta_2 X^2).$$

Note that these specifications are consistent with the identification conditions in terms of $\alpha_1 > 0$ and $\beta_2 \geq 0$. To meet the normalization that the hospital's bargaining power lies on $(0, 1)$, the hospital i's heterogeneous bargaining power is specified as

$$\gamma(Z_i) = \frac{\exp(\gamma_0 + Z_i'\gamma_1)}{1 + \exp(\gamma_0 + Z_i'\gamma_1)} \in (0, 1).$$

In the parametric model, $\alpha \equiv (\alpha_0, \alpha_1, \alpha_2, \alpha_3', \alpha_4')'$, $\beta \equiv (\beta_0, \beta_1, \beta_2)'$, and $\gamma \equiv (\gamma_0, \gamma_1')'$ are unknown parameter vectors.

The relevance of these specifications is reinforced based on the related empirical literature. First, besides the general arguments in section 2, the specification that each hospital's

ulation growth because the population growth may affect the marginal revenue in the treatment for patients with coronary heart disease. For example, [83] argue that the population affects the number of hospital entry, and an intensified competition shrinks the profit margin in the coronary heart disease market. Acknowledging this, the negotiated reimbursement rate may depend also on the population growth, not only on the current population. An alternative variable is the population age, which may affect the marginal cost of the treatment with coronary stents. As [84] find, there is relatively large impact of population ageing on the marginal treatment cost for patients with coronary stent disease in Australia, because the per-unit treatment costs may depend on the severity of the patient's illness and severity generally increases with age. However, we do not need such variables for the identification of parametric specifications. Even though we add them into our model, this does not affect the estimates of the parameters in the payment equation as well as the the counterfactual effect of bargaining power on the payment reduction because the payment equation does not include these variables given the additively separability of population growth in the model specification, which is consistent with the condition of nonparametric identification.

revenue depends on the aggregate demand X is supported by the closely related empirical literature including [78] in which the reimbursement rate transferred from insurers to hospitals depends on the hospitals' market demand since the hospital's revenue is mainly generated through the reimbursement from patient's insurer to hospital. Second, hospital's bargaining power may depend on hospitals' characteristics such as on the status of for-proft (or ownership) and teaching status ([81]).

We obtain the regression model under passive beliefs and symmetric beliefs, respectively, by substituting the parametric specifications into the optimal payment equation

$$\widetilde{P}_{t,i} = \frac{\overline{\gamma}_t \gamma_{t,i}}{1-\gamma_{t,i}} \left[c(X_t) - X_t c'(X_t) + \sum_{j=1}^{N} Q_{t,j}^2 u'(X_t, Z_{t,j}) \right] \\ + Q_{t,i} c'(X_t) - Q_{t,i}^2 u'(X_t, Z_{t,i}) + \eta_{t,i}, \tag{4.6}$$

$$\widetilde{P}_{t,i} = \frac{\overline{\gamma}_t \gamma_{t,i}}{1-\gamma_{t,i}} \left[c(X_t) - X_t c'(X_t) + N \sum_{j=1}^{N} Q_{t,j}^2 u'(X_t, Z_{t,j}) \right] \\ + Q_{t,i} c'(X_t) - N Q_{t,i}^2 u'(X_t, Z_{t,i}) + \eta_{t,i}, \tag{4.7}$$

where $\overline{\gamma}_t = \left(1 + \sum_{i=1}^{N} \frac{\gamma_{t,i}}{1-\gamma_{t,i}}\right)^{-1}$, and $\gamma_{t,i} \equiv \gamma(Z_{t,i})$. Clearly, the estimation procedures are similar between passive beliefs and symmetric beliefs, without loss of generality, we focus on the estimation with passive beliefs. Let $\lambda \equiv (\alpha, \beta, \gamma)$ with $\lambda_1 = \lambda / \{\alpha_0, \alpha_3\}$ and $\lambda_2 = \{\alpha_0, \alpha_3\}$. Under the conditional mean zero $\mathbf{E}(\eta_{t,i} | \boldsymbol{Q}_t, Z_{t,i}) = 0$, where $\boldsymbol{Q}_t = (Q_{t,i}, \cdots, Q_{t,N})'$, the system nonlinear least square estimator of λ_1 is obtained by

$$\widehat{\lambda}_1 = \underset{\lambda_1}{\text{argmin}}\, (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \eta_{t,i}^2 = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\widetilde{P}_{t,i} - P_{t,i} \right)^2 .$$

Similarly, based on the optimal quantity equation we obtain the second regression model $\xi_t = N c'(X_t) - \sum_{i=1}^{N} u(X_t, Z_{t,i}) - \sum_{i=1}^{N} Q_{t,i} u'(X_t, Z_{t,i})$. Under the assumption of conditional mean zero of ξ_t, we obtain the ordinary least square estimator $\widehat{\lambda}_2$ by plugging $\widehat{\lambda}_1$.[23]

To test for the equivalency of passive beliefs and symmetric beliefs, we use the nonnested test method proposed by [85] who generalize the seminal work [86] to a broad class of estimation methods including nonlinear least squares. Under the null hypothesis that model 1 associated with passive beliefs and model 2 associated with symmetric beliefs are asymptotically equivalent, the test statistic is $TS = \frac{\sqrt{TN}}{\widehat{\sigma}}\{Q^1 - Q^2\} \xrightarrow{d} N(0,1)$, where $Q^k = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}(\widehat{\eta}^k_{t,i})^2$ with $\widehat{\eta}^k_{t,i}$ for model $k \in \{1,2\}$, the standard error $\widehat{\sigma} = \left\{\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}[(\widehat{\eta}^1_{t,i})^4 + (\widehat{\eta}^2_{t,i})^4 - 2(\widehat{\eta}^1_{t,i})^2(\widehat{\eta}^2_{t,i})^2]\right\}^{1/2}$.

4.5.4 Empirical Results

Table 4.1 reports the estimation results of structural parameter under passive beliefs and symmetric beliefs, respectively. Based on these estimates, the statistic TS = -8.21 implies that at 1% significance level, we reject H_0 in favor of the hypothesis that passive beliefs fit the data better than symmetric beliefs. Moreover, the estimates under passive beliefs are consistent with the identification argument in section 3 as the marginal revenue function is strictly increasing and the cost function is strictly convex. Under symmetric beliefs, however, marginal revenue function is not necessarily increasing, depending on the total quantity of stents. More exactly, the marginal revenue is increasing only if the total quantity is very large (at least 3115).

According to the test results, we will focus on the analysis of estimation results under passive beliefs. First, marginal revenues for teaching hospitals are about $\$1$,000 more than nonteaching hospitals. The higher marginal revenue of teaching hospitals may arises from their competitive advantage due to better facilities and service, as they prefer to concentrate on research ([87]; [60]). As found by [80], the indicator of teaching hospital

[23] The feasible GLS method introduces additional estimation bias even though the feasible GLS is asymptotically efficient. Given the limited sample size, the OLS method seems more appropriate.

is positively and significantly related to market share. Using the dataset in 1997, [88] find that the average profit per coronary artery bypass graft surgery in New York state hospitals is \$2,000, which is larger than our estimates \$481.[24] The difference could be ascribed to the constantly increasing competition among hospitals over the past two decades. In addition, the marginal cost of manufacturer is about \$1,000 when the market size is not large, which is reasonable in the sense that it is smaller than the approximately average detrended price \$1,300.

Second, the hospital's bargaining powers is heterogeneous, depending on the public status of hospital. Specifically, the bargaining power of public hospitals is lower than that of private ones. This can be explained by the fact that public hospital concerns community interest more than profit ([60]), suggesting lower bargaining power due to less effort than private hospitals. In addition, the teaching status is insignificant in the bargaining power function which is consistent with [81].

Besides the heterogeneity, the bargaining power is very small for any characteristics of hospitals. To be specific, the bargaining power is 0.072 for private and nonteaching hospitals, 0.006 for private and teaching, 0.004 for public and nonteaching, and 0.0002 for public and teaching. Two common sources for the extremely small bargaining powers are market power of manufacturers and competition among hospitals. Furthermore, one important reason in the medical device industry is the substantial influences of doctor's brand-loyalty on the hospital's bargaining power because the primacy of physician preference in determining demand for stents has limited hospitals' bargaining ability in negotiation with suppliers. Therefore, the influence of doctor's brand-loyalty on the bargaining power may explain the policy of advocating gainsharing programs in cardiology and cardiac surgery by the U.S. Department of Health and Human Services Office of In-

[24] For the comparability, we use type-adjusted price to calculate the average profit per stent instead of type-month detrended price.

spector General, which will be analyzed in terms of the counterfactual results below.

Table 4.1: Estimates of structural parameters

	Parameters	Passive belief		Symmetric belief	
		Est.	Std.	Est.	Std.
Marginal Revenue	constant	991.2***	33.31	204.48***	8.244
	total quant.	-0.140	0.680	-0.218***	0.072
	total quant. $\times$ total quant.	0.0002***	0.00003	0.00007***	0.00001
	public	-73.02	100.3	-76.843***	25.464
	teaching	909.2***	71.77	180.797***	18.604
	public $\times$ total quant.	0.001	0.081	0.008	0.039
	teaching $\times$ total quant.	-0.647	0.618	-0.052	0.048
Cost	constant	1645.2***	57.321	3188.8***	0.544
	total quant.	963.8***	52.066	15.484***	1.960
	total quant. $\times$ total quant.	0.077***	0.014	0.325***	0.090
Bargaining	constant	-3.099***	0.029	-0.510***	0.154
	public	-2.650***	0.022	-7.251***	0.232
	teaching	-2.548	233.939	-2.456	50.477

Standard errors are bootstrapped 500 times. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.5.5 Counterfactual Analysis

The price of medical devices is often considered as a source of the increasing costs of healthcare ([89]). For example, the cost of hospital supplies and devices is estimated to account for 24 percent of the dramatic growth in inpatient hospital costs in nine states between 2001 and 2006 ([90]). Many interventions aim to lower these costs by increasing the bargaining power of hospitals in negotiation with suppliers. Since 2005 the U.S. Department of Health and Human Services Office of Inspector General has issued a handful of advisory opinions permitting additional gainsharing programs in cardiology and cardiac surgery, in which physicians receive cash payments for reducing hospital spending. The empirical evidence suggests that gainsharing program reduces the price of coronary stents, possibly as a result of the increased bargaining power of hospitals in negotiation with suppliers ([91]).[25]

To assess of the price effect of hospital's bargaining power, we will show price reduction obtained by continuously improving the bargaining power by the same amount (say, 0.01) for all hospitals. Increasing the bargaining power by the same amount seems more relevant than increasing it to the same level since it seems unrealistic for interventions to keep all hospitals with the same bargaining power, unless the bargaining power is irrelevant to hospitals' characteristics. Even if the bargaining power was homogeneous across hospitals, the magnitude of price changes might vary across hospitals because prices also depend on the heterogeneity of hospital revenue. More exactly, as equation (4.6) shows, average prices rely on the interaction of bargaining power and marginal revenue, suggesting the counterfactual result may depend on hospitals' characteristics through revenue functions.

For the purpose of evaluating the heterogeneous price effect of increasing bargaining power, we will calculate the average decrease of prices across hospital and over time as the bargaining power increases by the same amount, given each value of characteristics (public status and teaching status) of hospitals. Specifically, given the characteristic $Z = z$, we obtain the benchmark price $p^0 \equiv (TN)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \widehat{p}_{t,i}^0$, where the fitted price is defined as $\widehat{p}_{t,i}^0 \equiv \widehat{P}_{t,i}/Q_{t,i}$, and $\widehat{P}_{t,i}$ is the fitted value of payment $\widetilde{P}_{t,i}$. Suppose the bargaining power is increased J times. For each $j = 1, \cdots, J$, let $\overline{\gamma}_t^j = \left(1 + \sum_{i=1}^{N} \frac{\gamma_{t,i}^j}{1-\gamma_{t,i}^j}\right)^{-1}$, where $\gamma_{t,i}^j = \widehat{\gamma}_{t,i} + 0.01$, $\widehat{\gamma}_{t,i}$ is obtained by replacing $\gamma = (\gamma_0, \gamma_1)$ with $\widehat{\gamma} = (\widehat{\gamma}_0, \widehat{\gamma}_1)$ in $\gamma(Z_i)$. Let

$$\widehat{P}_{t,i}^j = \frac{\overline{\gamma}_t^j \gamma_{t,i}^j}{1 - \gamma_{t,i}^j} \left[\widehat{c}(X_t) - X_t \widehat{c}'(X_t) + \sum_{j=1}^{N} Q_{t,j}^2 \widehat{u}'(X_t, Z_{t,j})\right] + Q_{t,i} \widehat{c}'(X_t) - Q_{t,i}^2 \widehat{u}'(X_t, Z_{t,i}),$$

[25] In contrast to group purchasing organizations (GPOs) or merger among hospitals through which the collective bargaining power increases to increase aggregate profits by deceasing prices, one advantage of the gainsharing program lies in the fact that it can show show the way how the individual price (and hence profit) of each hospital change with its own bargaining power since in the former counterfactuals, it is not clear how to split the incremental aggregate profit among the group of hospitals.

where $\widehat{c}(X_t) = c(X_t;\widehat{\beta})$, $\widehat{c}'(X_t) = c'(X_t;\widehat{\beta})$, $\widehat{u}'(X_t) = u'(X_t;\widehat{\alpha})$, Then, we obtain the counterfactual prices $p^j \equiv (TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\widehat{p}_{t,i}^{j}$, where $\widehat{p}_{t,i}^{j} \equiv \widehat{P}_{t,i}^{j}/Q_{t,i}$, and hence the decrease of price is $p^{j,0} \equiv p^j - p^0$. Therefore, we obtain the series of the cut-down of prices $\{p^{j,0}\}_{j=1}^{J}$ for each of all four values of the binary vector Z.

Figure 4.4 shows the trend of price changes when the bargaining power increases continuously for each characteristic of hospitals and each of two beliefs. First, the price decreases as the bargaining power increases, for any hospital's characteristics and for either passive beliefs or symmetric beliefs. More importantly, under passive beliefs decreases of prices are more reasonable than under symmetric beliefs. As Figure 4.3 shows, the highest price is around $1,700, suggesting that the reasonable counterfactual price reduction should be less than $1,700. In Figure 4.4, price reductions under passive beliefs are less than $500, whereas price reductions under symmetric beliefs are mostly larger than $1,700. Therefore, those counterfactual results are not only consistent with results of belief tests, they also provide empirical support for the relevance of beliefs test because in the counterfactual analysis symmetric beliefs tend to lead to unreasonable and misleading policy implications.

Second, we focus counterfactual results under passive beliefs and compare the price effect of bargaining power between different hospitals' characteristics. In the upper panel of Figure 4.4, it indicates that private hospitals benefit more price reduction than public hospitals. One plausible explanation for the difference is due to higher bargaining power of private hospitals than public hospitals because public status is significantly negative in the bargaining power function. Note that price reduction also depends on the heterogeneous marginal revenue function in some nonlinear way, as Figure 4.4 shows, the overall effect of both bargaining power and heterogeneous marginal revenue determines the slightly nonlinear trend of price reductions as bargaining power increases. Similarly, in the lower panel of Figure 4.4, nonteaching hospitals benefit more price reduction than teaching hospitals.

Since the marginal revenue function is increasing in the teaching status, nonteaching hospitals might have more incentive to lower prices by increasing their bargaining powers in order to pursue the similar revenue with teaching hospitals. As such, bargaining power may have larger price effect for nonteaching hospitals.

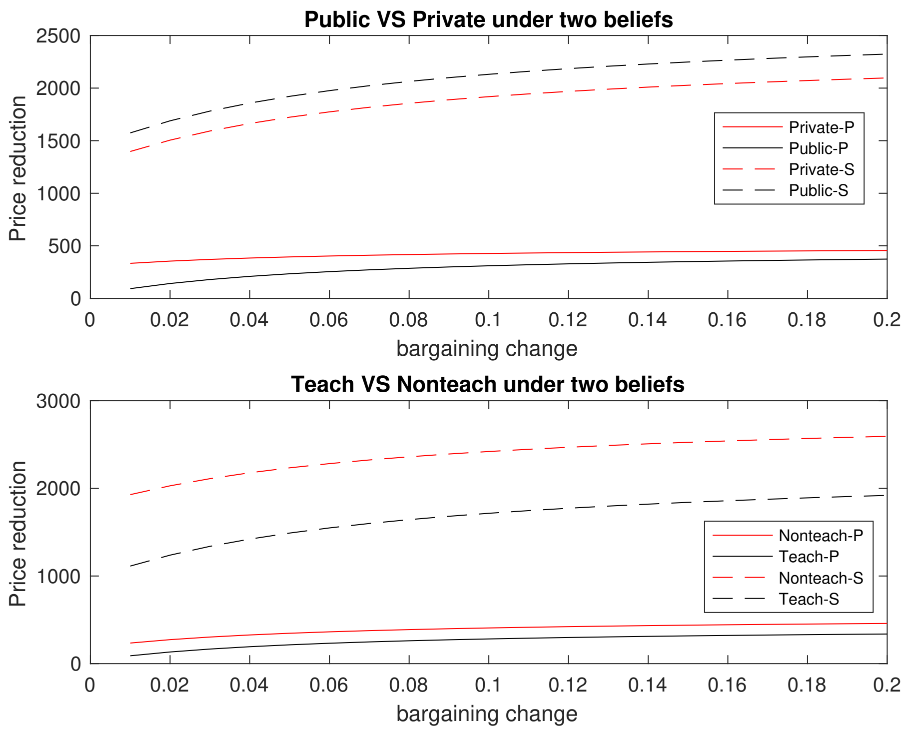

Figure 4.4: Price reductions under two beliefs in northeastern region

4.6 Concluding Remarks

In this paper, we have presented positive results in the identification of structural elements in the canonical model of contracting with externalities under passive beliefs and symmetric beliefs. Given the observed payments and quantities, the model structure consisting of the principal's cost function, the agent's payoff function, the bargaining power function, and the distribution of the random shocks of agents can be point identified under functional form restrictions on the density of shocks and on the large support of characteristics. The identification strategy we have used in this paper relies crucially on the structural one-to-one mapping between the quantity vector and the shock vector. We have proposed nonparametric estimators for the primitives, including the sieve estimator of cost function, the payoff function, and bargaining function, as well as the kernel estimator of the joint density of agents' shocks. Then, we have provided sufficient conditions to establish the rate of uniform convergence of these estimators.

We apply the model to analyze the dataset on the contracts of coronary stents between manufacturers and hospitals in the United States, and find that the model under passive beliefs fits the dataset better than symmetric beliefs. Further, the counterfactual analysis shows that the results with passive beliefs are more reasonable than with symmetric beliefs, which is consistent with our test results.

5. CONCLUSION

First, we relax the assumption of identity cost function in the model of FPCR menu and investigate how the performance of the optimal FPCR menu would be impacted by a convex cost function. We provide an important observation for the optimal FPCR menu when the agent's innate cost is uniformly distributed: the performance of the optimal FPCR contract relies crucially on the cost function: when the marginal cost (relative to marginal disutility) is large, the performance is very close to that of the fully optimal contract. On the other hand, if the marginal cost is small and firms' innate costs are dispersed then the performance of the optimal FPCR menu is arbitrarily close to a CR contract. Our result is in contrast with that of [1], where the FPCR menu captures at least three-fourths of the surplus that a fully optimal contract achieves. The main force that leads to the discrepancy is that under a convex cost function, the optimal cost-reducing effort exerted by an agent is strictly increasing in her innate cost and a cost function with higher marginal cost induces larger cost-reducing effort for given innate cost. By contrast, an identity cost function implies that the optimal cost-reducing effort is a constant, regardless of the innate cost of the agent. Our finding suggests that in designing an optimal FPCR contract it is important for the principal to take into account the cost structure of the agents: when the marginal cost of agents is large, the FPCR menu is especially preferable. When the marginal cost is small and agents' innate costs are dispersed, the menu is less appealing or even arbitrarily close to a CR contract.

Second, we provided a rigorous econometric analysis of the two-period FPCR contracts under both renegotiation and commitment, where we generalize the widely used identity cost function to a convex one and allow for a heterogeneous disutility function of effort. We proved that the model is nonparametrically identified if firms exert effort and the result of identification can be applied to a large class of contracts with and without incentives. If we include the firms without exerting effort, the model is semi nonparametrically identified. Based on the identification argument, we propose a feasible procedure to estimate the model primitives. Using the public transport procurement contracts in France, we found that cost function of firms are convex and the convexity has important implications for the welfare analysis: if the contract is switched from renegotiation to commitment, both taxpayers and firms would benefit and the major gains of social welfare would accrue to taxpayers.

Third, we have presented positive results in the identification of structural elements in the canonical model of contracting with externalities under passive beliefs and symmetric beliefs. Given the observed payments and quantities, the model structure consisting of the principal's cost function, the agent's payoff function, the bargaining power function, and the distribution of the random shocks of agents can be point identified under functional form restrictions on the density of shocks and on the large support of characteristics. The identification strategy we have used in this paper relies crucially on the structural one-to-one mapping between the quantity vector and the shock vector. We have proposed nonparametric estimators for the primitives, including the sieve estimator of cost function, the payoff function, and bargaining function, as well as the kernel estimator of the joint density of agents' shocks. Then, we have provided sufficient conditions to establish the rate of uniform convergence of these estimators. We apply the model to analyze the dataset on the contracts of coronary stents between manufacturers and hospitals in the United States, and find that the model under passive beliefs fits the dataset better than symmetric beliefs. Further, the counterfactual analysis shows that the results with passive beliefs are more reasonable than with symmetric beliefs, which is consistent with our test results.

REFERENCES

[1] W. P. Rogerson, “Simple menus of contracts in cost-based procurement and regulation,” *American Economic Review*, vol. 93, pp. 919–926, 2003.

[2] J.-J. Laffont and J. Tirole, “Using cost observation to regulate firms,” *Journal of Political Economy*, vol. 94, pp. 614–641, 1986.

[3] I. Perrigne and Q. Vuong, “Nonparametric identification of a contract model with adverse selection and moral hazard,” *Econometrica*, vol. 79, pp. 1499–1539, 2011.

[4] J.-J. Laffont and J. Tirole, *A Theory of Incentives in Procurement and Regulation*. MIT Press, Cambridge, 1993.

[5] I. Perrigne and Q. Vuong, “On the identification of the procurement model,” *Economics Letters*, vol. 114, pp. 9–11, 2012.

[6] P. Gagnepain and M. Ivaldi, “Incentive regulatory policies: the case of public transit systems in france,” *RAND Journal of Economics*, vol. 33, pp. 605–629, 2002.

[7] I. Segal, “Contracting with externalities,” *Quarterly Journal of Economics*, vol. 114, pp. 337–388, 1999.

[8] O. Hart and J. Tirole, “Vertical integration and market foreclosure,” *Brookings Papers on Economic Activity: Microeconomics*, vol. 1, pp. 205–286, 1990.

[9] R. P. McAfee and M. Schwartz, “Opportunism in multilateral vertical contracting: Nondiscrimination, exclusivity, and uniformity,” *American Economic Review*, vol. 84, pp. 210–230, 1994.

[10] E. B. Rasmusen, J. M. Ramseyer, and J. S. Wiley Jr, “Naked exclusion,” *American Economic Review*, vol. 81, pp. 1137–1145, 1991.

[11] I. R. Segal and M. D. Whinston, "Naked exclusion: comment," *American Economic Review*, vol. 90, pp. 296–309, 2000.

[12] M. L. Katz and C. Shapiro, "Technology adoption in the presence of network externalities," *Journal of Political Economy*, vol. 94, pp. 822–841, 1986.

[13] A. Raskovich, "Pivotal buyers and bargaining position," *Journal of Industrial Economics*, vol. 51, pp. 405–426, 2003.

[14] L. Y. Chu and D. E. Sappington, "Simple cost-sharing contracts," *American Economic Review*, vol. 97, pp. 419–428, 2007.

[15] J.-J. Laffont and D. Martimort, *The theory of incentives: The principal-agent problem.* Princeton University Press, Princeton, 2002.

[16] P. Bolton and M. Dewatripont, *Contract theory.* MIT press, Cambridge, 2005.

[17] P. Bajari and S. Tadelis, "Incentives versus transaction costs: A theory of procurement contracts," *RAND Journal of Economics*, vol. 32, pp. 387–407, 2001.

[18] P. Gagnepain, M. Ivaldi, and D. Martimort, "The cost of contract renegotiation: Evidence from the local public sector," *American Economic Review*, vol. 103, pp. 2352–83, 2013.

[19] A. V. Banerjee and E. Duflo, "Reputation effects and the limits of contracting: A study of the indian software industry," *Quarterly Journal of Economics*, vol. 115, pp. 989–1017, 2000.

[20] S. M. Schennach and Y. Hu, "Nonparametric identification and semiparametric estimation of classical measurement error models without side information," *Journal of the American Statistical Association*, vol. 108, pp. 177–186, 2013.

[21] Y. Hu, "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, vol. 144, pp. 27–61, 2008.

[22] Y. Hu and S. M. Schennach, "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, vol. 76, pp. 195–216, 2008.

[23] M. Dewatripont, "Renegotiation and information revelation over time: The case of optimal labor contracts," *Quarterly Journal of Economics*, vol. 104, pp. 589–619, 1989.

[24] J.-J. Laffont and J. Tirole, "Adverse selection and renegotiation in procurement," *Review of Economic Studies*, vol. 57, pp. 597–625, 1990.

[25] Baron and Myerson, "Regulating a monopolist with unknown costs," *Econometrica*, vol. 50, pp. 911–930, 1982.

[26] D. P. Baron, "Regulation and legislative choice," *RAND Journal of Economics*, vol. 19, pp. 467–477, 1988.

[27] A. Faure-Grimaud and D. Martimort, "Regulatory inertia," *Rand Journal of Economics*, vol. 34, pp. 413–437, 2003.

[28] D. F. Garrett, "Robustness of simple menus of contracts in cost-based procurement," *Games and Economic Behavior*, vol. 87, pp. 631–641, 2014.

[29] O. D. Hart and J. Tirole, "Contract renegotiation and coasian dynamics," *Review of Economic Studies*, vol. 55, pp. 509–540, 1988.

[30] T. Li, "Robust and consistent estimation of nonlinear errors-in-variables models," *Journal of Econometrics*, vol. 110, pp. 1–26, 2002.

[31] S. Cicala, "When does regulation distort costs? lessons from fuel procurement in us electricity generation," *American Economic Review*, vol. 105, pp. 411–444, 2015.

[32] S. Bhattacharya and J. L. Guasch, "Heterogeneity, tournaments, and hierarchies," *Journal of Political Economy*, vol. 96, pp. 867–881, 1988.

[33] E. Guerre, I. Perrigne, and Q. Vuong, "Nonparametric Identification of Risk Aversion in First-price Auctions under Exclusion Restrictions," *Econometrica*, vol. 77, pp. 1193–1227, 2009.

[34] E. Guerre, I. Perrigne, and Q. Vuong, "Optimal nonparametric estimation of first-price auctions," *Econometrica*, vol. 68, pp. 525–574, 2000.

[35] T. Magnac and D. Thesmar, "Identifying dynamic discrete decision processes," *Econometrica*, vol. 70, pp. 801–816, 2002.

[36] W. Roy and A. Yvrande-Billon, "Ownership, contractual practices and technical efficiency: The case of urban public transport in france," *Journal of Transport Economics and Policy*, vol. 41, pp. 257–282, 2007.

[37] J. P. Kalt and M. A. Zupan, "Capture and ideology in the economic theory of politics," *American Economic Review*, vol. 74, pp. 279–300, 1984.

[38] C. L. Ballard, J. B. Shoven, and J. Whalley, "General equilibrium computations of the marginal welfare costs of taxes in the united states," *American Economic Review*, vol. 75, pp. 128–138, 1985.

[39] X. Chen and D. Pouzo, "Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals," *Econometrica*, vol. 80, pp. 277–321, 2012.

[40] R. L. Matzkin, "Identification in nonparametric simultaneous equations models," *Econometrica*, vol. 76, pp. 945–978, 2008.

[41] C. Ai and X. Chen, "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, vol. 71, pp. 1795–1843, 2003.

[42] R. L. Matzkin, “Estimation of nonparametric models with simultaneity,” *Econometrica*, vol. 83, pp. 1–66, 2015.

[43] A. Chesher, “Identification in nonseparable models,” *Econometrica*, vol. 71, pp. 1405–1441, 2003.

[44] G. W. Imbens and W. K. Newey, “Identification and estimation of triangular simultaneous equations models without additivity,” *Econometrica*, vol. 77, pp. 1481–1512, 2009.

[45] A. Torgovitsky, “Identification of nonseparable models using instruments with small support,” *Econometrica*, vol. 83, pp. 1185–1197, 2015.

[46] X. D’Haultfœuille and P. Février, “Identification of nonseparable triangular models with discrete instruments,” *Econometrica*, vol. 83, pp. 1199–1210, 2015.

[47] R. L. Matzkin, “Nonparametric estimation of nonadditive random functions,” *Econometrica*, vol. 71, pp. 1339–1375, 2003.

[48] S. Hoderlein and E. Mammen, “Identification of marginal effects in nonseparable models without monotonicity,” *Econometrica*, vol. 75, pp. 1513–1518, 2007.

[49] T. Li, I. Perrigne, and Q. Vuong, “Structural estimation of the affiliated private value auction model,” *RAND Journal of Economics*, vol. 33, pp. 171–193, 2002.

[50] A. Merlo and X. Tang, “Identification and estimation of stochastic bargaining models,” *Econometrica*, vol. 80, pp. 1563–1604, 2012.

[51] E. D. Gould, B. P. Pashigian, and C. J. Prendergast, “Contracts, externalities, and incentives in shopping malls,” *Review of Economics and Statistics*, vol. 87, pp. 411–422, 2005.

[52] S. Ferrari and F. Verboven, “Vertical control of a distribution networkŮ: an empirical analysis of magazines,” *RAND Journal of Economics*, vol. 43, pp. 26–50, 2012.

[53] J. Asker, "Diagnosing foreclosure due to exclusive dealing," *Journal of Industrial Economics*, vol. 64, pp. 375–410, 2016.

[54] S. T. Berry, "Estimating discrete-choice models of product differentiation," *RAND Journal of Economics*, vol. 25, pp. 242–262, 1994.

[55] S. Berry, J. Levinsohn, and A. Pakes, "Automobile prices in market equilibrium," *Econometrica*, vol. 63, pp. 841–890, 1995.

[56] A. Nevo, "Empirical models of consumer behavior," *Annual Review of Economics*, vol. 3, pp. 51–75, 2011.

[57] C. M. Kahn and D. Mookherjee, "Competition and incentives with nonexclusive contracts," *RAND Journal of Economics*, vol. 29, pp. 443–465, 1998.

[58] A. Bisin and D. Guaitoli, "Moral hazard and nonexclusive contracts," *RAND Journal of Economics*, vol. 35, pp. 306–328, 2004.

[59] R. Town and G. Vistnes, "Hospital competition in hmo networks," *Journal of Health Economics*, vol. 20, pp. 733–753, 2001.

[60] M. Gaynor and W. B. Vogt, "Competition among hospitals," *RAND Journal of Economics*, vol. 34, pp. 764–785, 2003.

[61] A. Sweeting, "The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria," *RAND Journal of Economics*, vol. 40, pp. 710–742, 2009.

[62] A. De Paula and X. Tang, "Inference of signs of interaction effects in simultaneous games with incomplete information," *Econometrica*, vol. 80, pp. 143–172, 2012.

[63] H. Horn and A. Wolinsky, "Bilateral monopolies and incentives for merger," *RAND Journal of Economics*, vol. 19, pp. 408–419, 1988.

[64] A. Arya and B. Mittendorf, "Disclosure standards for vertical contracts," *Rand Journal of Economics*, vol. 42, pp. 595–617, 2011.

[65] P. Rey and T. Vergé, "Bilateral control with vertical contracts," *RAND Journal of Economics*, vol. 35, pp. 728–746, 2004.

[66] K. Binmore, A. Rubinstein, and A. Wolinsky, "The nash bargaining solution in economic modelling," *RAND Journal of Economics*, vol. 17, pp. 176–188, 1986.

[67] C. Milliou and E. Petrakis, "Upstream horizontal mergers, vertical contracts, and bargaining," *International Journal of Industrial Organization*, vol. 25, pp. 963–987, 2007.

[68] C. De Fontenay and J. S. Gans, "Vertical integration in the presence of upstream competition," *RAND Journal of Economics*, vol. 36, pp. 544–572, 2005.

[69] G. S. Crawford and A. Yurukoglu, "The welfare effects of bundling in multichannel television markets," *American Economic Review*, vol. 102, pp. 643–685, 2012.

[70] A. Rubinstein, "Perfect equilibrium in a bargaining model," *Econometrica*, vol. 50, pp. 97–109, 1982.

[71] Y. Chen and M. H. Riordan, "Vertical integration, exclusive dealing, and expost cartelization," *RAND Journal of Economics*, vol. 38, pp. 1–21, 2007.

[72] M. Pagnozzi and S. Piccolo, "Vertical separation with private contracts," *Economic Journal*, vol. 122, pp. 173–207, 2012.

[73] S.-Y. Chou, M. E. Deily, S. Li, and Y. Lu, "Competition and the impact of online hospital report cards," *Journal of Health Economics*, vol. 34, pp. 42–58, 2014.

[74] R. W. Hahn, K. B. Klovers, and H. J. Singer, "The need for greater price transparency in the medical device industry: an economic analysis," *Health Affairs*, vol. 27, pp. 1554–1559, 2008.

[75] J. C. Robinson, “Variation in hospital costs, payments, and profitabilty for cardiac valve replacement surgery,” *Health Services Research*, vol. 46, pp. 1928–1945, 2011.

[76] M. Grennan, “Price discrimination and bargaining: Empirical evidence from medical devices,” *American Economic Reivew*, vol. 103, pp. 145–177, 2013.

[77] J. E. Kralewski, R. Feldrnan, B. Dowd, and J. Shapiro, “Strategies employed by hmos to achieve hospital discounts: A case study of seven hmos.,” *Health Care Management Review*, vol. 16, pp. 9–16, 1991.

[78] G. Gowrisankaran, A. Nevo, and R. Town, “Mergers when prices are negotiated: Evidence from the hospital industry,” *American Economic Review*, vol. 105, pp. 172–203, 2015.

[79] J. M. Brooks, A. Dor, and H. S. Wong, “Hospital-insurer bargaining: an empirical investigation of appendectomy pricing,” *Journal of Health Economics*, vol. 16, pp. 417–434, 1997.

[80] K. Ho, “Insurer-provider networks in the medical care market,” *American Economic Review*, vol. 99, pp. 393–430, 2009.

[81] M. S. Lewis and K. E. Pflum, “Diagnosing hospital system bargaining power in managed care networks,” *American Economic Journal: Economic Policy*, vol. 7, pp. 243–274, 2015.

[82] C. Shih and E. Berliner, “Diffusion of new technology and payment policies: coronary stents,” *Health Affairs*, vol. 27, pp. 1566–1576, 2008.

[83] S. Li and A. Dor, “How do hospitals respond to market entry? evidence from a deregulated market for cardiac revascularization,” *Health Economics*, vol. 24, pp. 990–1008, 2015.

[84] L. Thurecht, A. Walker, A. Harding, and J. Pearse, "The inverse care law, population ageing and the hospital system: a distributional analysis," *Economic Papers: A Journal of Applied Economics and Policy*, vol. 24, pp. 1–17, 2005.

[85] D. Rivers and Q. Vuong, "Model selection tests for nonlinear dynamic models," *Econometrics Journal*, vol. 5, pp. 1–39, 2002.

[86] Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, vol. 57, pp. 307–333, 1989.

[87] M. Noether, "Competition among hospitals," *Journal of Health Economics*, vol. 7, pp. 259–284, 1988.

[88] R. S. Huckman, "Hospital integration and vertical consolidation: An analysis of acquisitions in new york state," *Journal of Health Economics*, vol. 25, pp. 58–80, 2006.

[89] S. P. Keehan, A. M. Sisko, C. J. Truffer, J. A. Poisal, G. A. Cuckler, A. J. Madison, J. M. Lizonitz, and S. D. Smith, "National health spending projections through 2020: economic recovery and reform drive faster spending growth," *Health Affairs*, vol. 36, pp. 10–1377, 2011.

[90] J. L. K. Maeda, S. O. Raetzman, and B. S. Friedman, "What hospital inpatient services contributed the most to the 2001–2006 growth in the cost per case?," *Health Services Research*, vol. 47, pp. 1814–1835, 2012.

[91] J. D. Ketcham and M. F. Furukawa, "Hospital-physician gainsharing in cardiology," *Health Affairs*, vol. 27, pp. 803–812, 2008.

[92] Q. Li and J. S. Racine, *Nonparametric econometrics: theory and practice*. Princeton University Press, Princeton, 2007.

[93] E. Masry, "Multivariate local polynomial regression for time series: uniform strong consistency and rates," *Journal of Time Series Analysis*, vol. 17, pp. 571–599, 1996.

APPENDIX A

SIMPLE MENUS OF COST-BASED CONTRACTS WITH CONVEX COST FUNCTIONS

A.1 Proof of Lemma 1

Let $\Delta \equiv \bar{\theta} - \underline{\theta}$. Suppose that $\theta \in [\underline{\theta}, \bar{\theta}]$ is the cut-off type. By plugging the optimal effort level $e^*(\theta) = \beta\theta/(\beta+1)$, the principal's expected cost for the project is calculated as

$$C(\theta) = \Big[H(\theta - e^*(\theta)) + \psi(e^*(\theta))\Big]F(\theta) + \int_{\theta}^{\bar{\theta}} H(\theta)dF(\theta) = \frac{\beta\theta^2(\theta - \underline{\theta})}{\Delta(\beta+1)} + \frac{\beta(\bar{\theta}^3 - \theta^3)}{3\Delta},$$

with the first-order condition being:

$$C'(\theta) \quad = \quad \frac{\beta\theta}{\Delta(\beta+1)}\big[(2-\beta)\theta - 2\underline{\theta}\big].$$

The optimal effort level θ^* is given by $C'(\theta^*) = 0$,

$$\theta^* = 0; \text{ or } \theta^* = \frac{2\underline{\theta}}{2-\beta}, \beta \neq 2.$$

When $\beta \geq 2$ or $\gamma \leq 2/(2-\beta)$, $\theta^* = 2\underline{\theta}/(2-\beta) \notin (\underline{\theta}, \bar{\theta})$. Thus the cut-off type is not an interior point of the support $[\underline{\theta}, \bar{\theta}]$ and it is necessary to compare $C(\underline{\theta})$ with $C(\bar{\theta})$ to determine the optimal FPCR contract. Considering that

$$C(\underline{\theta}) \quad = \quad \frac{\beta}{3}(\bar{\theta}^2 + \bar{\theta}\underline{\theta} + \underline{\theta}^2); \;\; C(\bar{\theta}) = \frac{\beta}{\beta+1}\bar{\theta}^2.$$

It is straightforward to verify that the ratio $C(\underline{\theta})/C(\bar{\theta})$ is greater than one whenever $\gamma \leq 2/(2-\beta)$. Thus, cut-off type is $\bar{\theta}$ and the principal sets the fixed-price to be $b = H(\bar{\theta} - e^*(\bar{\theta})) + \psi(e^*(\bar{\theta})) = \beta\bar{\theta}^2/(\beta+1)$ and all the firms choose the FP contract.

When $0 < \beta < 2(\gamma-1)/\gamma$, $\theta^* = 2\underline{\theta}/(2-\beta) \in (\underline{\theta}, \bar{\theta})$. The second-order derivative of $C''(\theta^*) = 2\beta\underline{\theta}/[(\beta+1)\Delta] > 0$. Hence θ^* is an interior solution of the principal's cost minimization problem.

A.2 Proof of Lemma 2

For the optimal contract in Laffont and Tirole (1986), the principal solve the following optimization problem:

$$\min_{t(\theta),e(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} t(\theta) dF(\theta) \tag{A.1}$$

$$\text{s.t. } U(\theta|\theta) \geq 0$$

$$U(\theta|\theta) \geq U(\hat{\theta}|\theta)$$

$$\theta - e(\hat{\theta}|\theta) = \hat{\theta} - e(\hat{\theta}), \tag{A.2}$$

where $U(\hat{\theta}|\theta) \equiv t(\hat{\theta}) - H\left(\theta - e(\hat{\theta}|\theta)\right) - \psi(e(\hat{\theta}|\theta))$ and (A.2) is due to the strictly monotonicity of $H(\cdot)$. According to Laffont and Tirole (1986, 1993), the level of cost reduction chosen by the firm $e(\theta)$ is given by:

$$e(\theta) = \arg\min_{e\geq 0} \left\{ H(\theta - e(\theta)) + \psi(e(\theta)) + \psi'(e(\theta))\frac{F(\theta)}{f(\theta)} \right\}. \tag{A.3}$$

with the first-order condition being

$$H'\left(\theta - e(\theta)\right) - \psi'(e(\theta)) - \psi''(e(\theta))\frac{F(\theta)}{f(\theta)} = 0. \tag{A.4}$$

Plugging the parametric form of $H(\cdot)$ and $\psi(\cdot)$ to the equation above, we have

$$\beta(\theta - e) - e - (\theta - \underline{\theta}) = 0 \Rightarrow (1+\beta)e = (\beta - 1)\theta + \underline{\theta}. \tag{A.5}$$

Case 1. $(\beta - 1)\theta + \underline{\theta} \leq 0$, i.e., $0 < \beta < 1$ and $(1-\beta)\theta \geq \underline{\theta}$ then $e = 0$.

Case 2. $(\beta - 1)\theta + \underline{\theta} \geq 0$, i.e., $\beta \geq 1$ or $\beta < 1$ and $(1-\beta)\theta \leq \underline{\theta}$, then $e = \frac{(\beta-1)\theta+\underline{\theta}}{\beta+1}$.

A.3 Proof of Theorem 1

We first write down the surplus under the FPCR menu G_F,

$$\begin{aligned} G_F &\equiv C(\underline{\theta}) - C(\theta^*) = \int_{\underline{\theta}}^{\bar{\theta}} H(\theta)dF(\theta) - \Big[H(\theta^* - e^*(\theta^*)) + \psi(e^*(\theta^*))\Big]F(\theta^*) \\ &\quad - \int_{\theta^*}^{\bar{\theta}} H(\theta)dF(\theta) \\ &= \begin{cases} \frac{\beta}{3}\left(\frac{\beta-2}{\beta+1}\gamma^2 + \gamma + 1\right)\underline{\theta}^2, & 2(\gamma-1)/\gamma \leq \beta, \\ -\frac{\beta^3(\beta-3)}{3(\gamma-1)(\beta-2)^2(\beta+1)}\underline{\theta}^2, & 0 < \beta < 2(1-\gamma)/\gamma, \end{cases} \end{aligned} \tag{A.6}$$

Next, G_O can be obtained according to its definition:

$$G_O = \begin{cases} \int_{\underline{\theta}}^{\bar{\theta}} H(\theta)dF(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} \left\{H(\theta - e^*(\theta)) + \psi(e^*(\theta)) + \psi'(e^*(\theta))\frac{F(\theta)}{f(\theta)}\right\} dF(\theta), \\ 0 < \beta \leq (\gamma-1)/\gamma, \\ \int_{\underline{\theta}}^{\bar{\theta}} H(\theta)dF(\theta) - \int_{\underline{\theta}}^{\underline{\theta}/(1-\beta)} \left\{H(\theta - e^*(\theta)) + \psi(e^*(\theta)) + \psi'(e^*(\theta))\frac{F(\theta)}{f(\theta)}\right\} dF(\theta) \\ \qquad - \int_{\underline{\theta}/(1-\beta)}^{\bar{\theta}} H(\theta)dF(\theta), \ \beta > (\gamma-1)/\gamma. \end{cases}$$

Based on the optimal effort level in Lemma 2, we discuss G_O in two cases.

Case 1: $0 < \beta < 1$ and $(1-\beta)\bar{\theta} \geq \underline{\theta}$, i.e., $0 < \beta \leq (\gamma - 1)/\gamma$.

$$\begin{aligned}
G_O &= \int_{\underline{\theta}}^{\bar{\theta}} H(\theta)dF(\theta) - \int_{\underline{\theta}}^{\underline{\theta}/(1-\beta)} \left\{ H(\theta - e^*(\theta)) + \psi(e^*(\theta)) + \psi'(e^*(\theta))\frac{F(\theta)}{f(\theta)} \right\} dF(\theta) \\
&\qquad - \int_{\underline{\theta}/(1-\beta)}^{\bar{\theta}} H(\theta)dF(\theta) \\
&= \int_{\underline{\theta}}^{\underline{\theta}/(1-\beta)} \left\{ H(\theta) - H(\theta - e^*(\theta)) - \psi(e^*(\theta)) - \psi'(e^*(\theta))\frac{F(\theta)}{f(\theta)} \right\} dF(\theta) \\
&= \frac{\beta^3\underline{\theta}^3}{3(1-\beta^2)}\frac{1}{\bar{\theta}-\underline{\theta}} = \frac{\beta^3\underline{\theta}^2}{3(1-\beta^2)}\frac{1}{\gamma-1}. \qquad \text{(A.7)}
\end{aligned}$$

Case 2: $0 < \beta < 1$ and $(1-\beta)\bar{\theta} \leq \underline{\theta}$, or $\beta > 1$; i.e., $\beta > (\gamma - 1)/\gamma$.

$$\begin{aligned}
G_O &= \int_{\underline{\theta}}^{\bar{\theta}} H(\theta)dF(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} \left\{ H(\theta - e^*(\theta)) + \psi(e^*(\theta)) + \psi'(e^*(\theta))\frac{F(\theta)}{f(\theta)} \right\} dF(\theta) \\
&= \int_{\underline{\theta}}^{\bar{\theta}} \left\{ \beta\theta^2 - \beta\left(\frac{2\theta-\underline{\theta}}{1+\beta}\right)^2 - \left(\frac{(\beta-1)\theta+\underline{\theta}}{1+\beta}\right)^2 - 2\frac{(\beta-1)\theta+\underline{\theta}}{1+\beta}(\theta-\underline{\theta}) \right\} dF(\theta) \\
&= \frac{[(\beta-1)\bar{\theta}+\underline{\theta}]^3 - (\beta\underline{\theta})^3}{3(\beta^2-1)}\frac{1}{\bar{\theta}-\underline{\theta}} \\
&= \frac{1}{3(1+\beta)}\left\{(\beta-1)^2\bar{\theta}^2 + (\beta+2)(\beta-1)\bar{\theta}\underline{\theta} + (\beta^2+\beta+1)\underline{\theta}^2\right\} \\
&= \frac{(\beta-1)^2\gamma^2 + (\beta^2+\beta-2)\gamma + (\beta^2+\beta+1)}{3(1+\beta)}\underline{\theta}^2 \qquad \text{(A.8)}
\end{aligned}$$

The results in (A.7) and (A.8) allow us to obtain G_F/G_O as follows.

Case 1: $0 < \beta \leq 1 - 1/\gamma$,

$$\frac{G_F}{G_O} = \frac{(\beta-1)(\beta-3)}{(\beta-2)^2} = 1 - \frac{1}{(\beta-2)^2}. \qquad \text{(A.9)}$$

Case 2: $1 - 1/\gamma < \beta < 2(1 - 1/\gamma)$,

$$\frac{G_F}{G_O} = -\frac{\beta^3(\beta-3)}{(\gamma-1)(\beta-2)^2[(\beta-1)^2\gamma^2 + (\beta^2+\beta-2)\gamma + (\beta^2+\beta+1)]}. \qquad \text{(A.10)}$$

Case 3: $2(1-1/\gamma) \le \beta$,

$$\frac{G_F}{G_O} = 1 - \frac{\gamma^2+1-2\gamma}{(\beta-1)^2\gamma^2+(\beta^2+\beta-2)\gamma+(\beta^2+\beta+1)}. \tag{A.11}$$

It is easy to show that the ratio G_F/G_O is strictly decreasing in β for case 1 and $G_F/G_O < 3/4$ for any β in this case. On the other hand, the denominator of the second term on the right-hand-side of (A.11) is increasing in β and equals zero when $\beta = 2(\gamma-1)/\gamma$. As a result,

$$\frac{d}{d\beta}\frac{G_F}{G_O} > 0, \text{ if } 2(1-1/\gamma) \le \beta.$$

In addition, it is straightforward to verify that the limit of G_F/G_O is 1 as β approach infinity, i.e, $\lim_{\beta\to\infty}\frac{G_F}{G_O} = 1$.

Parts (i) and (iii) of Theorem 1 state that for any given $\gamma > 1$ the function G_F/G_O is decreasing in $\beta \in (0, 1-1/\gamma]$ and increasing in $\beta \in [2-2/\gamma), \infty)$. The continuity of G_F/G_O in γ implies there must be a minimum in the interval $\beta \in (1-1/\gamma, 2-2/\gamma)$.

A.4 Proof of Theorem 2

For any given $\beta \in (1-1/\gamma, 2-2/\gamma) \subset (0,2)$, we consider the ratio G_F/G_O:

$$\begin{aligned}\frac{G_F}{G_O} &= \frac{\beta^3(3-\beta)}{(\gamma-1)(\beta-2)^2[(\beta-1)^2\gamma^2+(\beta^2+\beta-2)\gamma+(\beta^2+\beta+1)]}\\ &= \frac{\beta^3(3-\beta)}{(\beta-2)^2}\frac{1}{(\gamma-1)[(\beta-1)^2\gamma^2+(\beta^2+\beta-2)\gamma+(\beta^2+\beta+1)]},\end{aligned}$$

where the first term $\beta^3(3-\beta)/(\beta-2)^2$ is a positive constant. Thus we focus on the denominator of the second term, denoted by $\xi(\gamma,\beta)$. For a given $\beta \in (1-1/\gamma, 2-2/\gamma)$,

$$\begin{aligned}\frac{d\xi(\gamma,\beta)}{d\gamma} &= 3\gamma^2(\beta-1)^2 - 2\gamma(\beta-1)^2 + 2\gamma(\beta^2+\beta-2) + (\beta^2+\beta-2) + (\beta^2+\beta+1) \\ &= 3\gamma^2\beta^2 + (6\gamma-6\gamma^2)\beta + 3\gamma^2 - 6\gamma + 3.\end{aligned}$$

It is easy to show that for any $\beta > 1-1/\gamma$, the quadratic form of β is positive, i.e., $d\xi(\gamma,\beta)/d\gamma > 0$ for any $\beta \in (1-1/\gamma, 2-2/\gamma)$. This completes the proof that G_F/G_O is strictly decreasing in γ for any $\beta \in (1-1/\gamma, 2-2/\gamma)$.

APPENDIX B

ECONOMETRICS OF MULTI-PERIOD SIMPLE CONTRACTS

B.1 Proofs of Main Results

Proof of Proposition 2. Before proving Proposition 2, note that many of the theoretical results under commitment carry over to the dynamic setting as below. For example, in both periods $\psi'(e^*(\theta)) = H'(\theta - e^*(\theta))$, and $0 < e^{*'}(\theta) < 1$. The proof below follows the arguments in [18]. Let $\Theta_G \equiv [\underline{\theta}, \theta_1^*]$, $\Theta_I \equiv (\theta_1^*, \theta_2^*]$, and $\Theta_B \equiv (\theta_2^*, \bar{\theta}]$. Denote respectively by $C_1^0 = (b_1, b_2^0)$, $C_2^0 = (H(\theta), b_3^0)$ and $C_3^0 = (H(\theta), H(\theta))$ the firm's payments under each scenario Θ_G, Θ_I, Θ_B, and by $C^0 = (b_1, b_2^0, b_3^0)$ the overall menu of fixed prices. Denote $\widetilde{R} = (\widetilde{C}_2, \widetilde{C}_3) = (\widetilde{b}_2, \widetilde{b}_3)$ as a subsidy profile offered at the renegotiation stage following an initial offer C^0 and

$$\tilde{b}_2 \geq b_2^0 \text{ and } \tilde{b}_3 \geq b_3^0. \tag{B.1}$$

Lemma 10. *(Renegotiation-proof) There is no loss of generality in restricting the analysis to contracts of the form $C = (b_1, R)$ that come unchanged through the renegotiation process, such that $R = (b_2, b_3)$ maximizes the principal?s second period welfare subject to the following acceptance conditions:*

$$\tilde{b}_2 \geq b_2 \textit{ and } \tilde{b}_3 \geq b_3. \tag{B.2}$$

Proof of Lemma 10: For any initial contract C^0 and consider a renegotiated offer $\widetilde{R} = (\widetilde{b}_2, \widetilde{b}_3)$ that satisfies (B.1). Given the the agent's conjecture about the renegotiated offer $R=(b_2,b_3)$, the principal ?s expected welfare for date 2 becomes[1]

$$\begin{aligned} SW_2(C^0,\widetilde{R},R) &= \int_{\underline{\theta}}^{\theta_1^*}\Big(S-(1+\lambda)\widetilde{b}_2+\alpha(\widetilde{b}_2-H(\theta-e^*(\theta))-\psi(e^*(\theta)))\Big)\,dF(\theta) \\ &+ \int_{\theta_1^*}^{\theta_2^*}\Big(S-(1+\lambda)\widetilde{b}_3+\alpha(\widetilde{b}_3-H(\theta-e^*(\theta))-\psi(e^*(\theta)))\Big)\,dF(\theta) \\ &+ \int_{\theta_2^*}^{\bar{\theta}}(S-(1+\lambda)H(\theta))\,f(\theta)d\theta \end{aligned} \tag{B.3}$$

Then the renegotiated offers $R=(b_2,b_3)$ must solve

$$R = \underset{\widetilde{R}}{argmax}\, SW_2(C^0,\widetilde{R},R) \quad \text{subject to (B.1).} \tag{$\mathbf{R}^0$}$$

Due to the arbitrary C^0, it is easy to obtain that R also solves the following problem

$$R = \underset{\widetilde{R}}{argmax}\, SW_2(C\equiv(b_1,R),\widetilde{R},R) \quad \text{subject to (B.2).} \tag{$\mathbf{R}$}$$

This completes the proof of Lemma 10.

Let us now characterize renegotiation-proof allocations by solving the problem **R**.

Lemma 11. *A first-period menu of contracts $C=(b_1,b_2,b_3)$ is renegotiation-proof if and only if the following two conditions hold:*

$$\theta_2^*(b_3) \geq \theta_1^*(b_1,b_2,b_3) \tag{B.4}$$

$$\left(1-\frac{\alpha}{1+\lambda}\right)[F(\theta_2^*)-F(\theta_1^*)] = \frac{[H(\theta_2^*)-b_3]f(\theta_2^*)}{H'(\theta_2^*-e^*(\theta_2^*))} \tag{B.5}$$

Condition (B.4) guarantees that the interval Θ_I is non-empty.

Proof of Lemma 11. First note that the assumption $\alpha<1+\lambda$ implies that the maximum

[1]Note that in $SW_2(C^0,\widetilde{R},R)$, $\theta_1^*=\theta_1^*(b_1,b_2,b_3)$ and $\theta_2^*=\theta_2^*(\widetilde{b}_3)$ not $\theta_2^*(b_3)$.

of the integral in (B.3) requires that (B.2) is binding. Assume that (**R**) is strictly quasi-concave in $\widetilde{b}_3$. The first-order condition of the optimization problem (**R**) with respect to $\widetilde{b}_3$ at $\widetilde{b}_3 = b_3$ is

$$\begin{aligned}
0 &= \frac{d\theta_2^*}{db_3}\left\{S-(1+\lambda)b_3+\alpha[b_3-H(\theta_2^*-e^*(\theta_2^*))-\psi(e^*(\theta_2^*))]\right\}f(\theta_2^*) \\
&\quad+\int_{\theta_1^*}^{\theta_2^*}(\alpha-1-\lambda)f(\theta)d\theta-\frac{d\theta_2^*}{db_3}[S-(1+\lambda)H(\theta_2^*)]f(\theta_2^*) \\
&= \frac{d\theta_2^*}{db_3}(1+\lambda)[H(\theta_2^*)-b_3]f(\theta_2^*)+\int_{\theta_1^*}^{\theta_2^*}(\alpha-1-\lambda)f(\theta)d\theta,
\end{aligned}$$

where $b_3 = H(\theta_2^* - e^*(\theta_2^*)) + \psi(e^*(\theta_2^*))$ will be proved in (B.7) below. Note that $1 = [H'(\theta_2^* - e^*(\theta_2^*))(1-e^{*'}(\theta_2^*)) + \psi'(e^*(\theta_2^*))e^{*'}(\theta_2^*)]d\theta_2^*/db_3 = H'(\theta_2^* - e^*(\theta_2^*))d\theta_2^*/db_3$, we obtain

$$(1+\lambda-\alpha)[F(\theta_2^*)-F(\theta_1^*)] = \frac{(1+\lambda)[H(\theta_2^*)-b_3]f(\theta_2^*)}{H'(\theta_2^*-e^*(\theta_2^*))},$$

which completes the proof of Lemma 11.

Define now the principal's intertemporal welfare when offering $C = (b_1, b_2, b_3)$ as[2]

$$\begin{aligned}
SW(C) &= \int_{\underline{\theta}}^{\theta_1^*}\{S-(1+\lambda)(rb_1+(1-r)b_2)+\alpha[rb_1+(1-r)b_2 \\
&\quad -H(\theta-e^*(\theta))-\psi(e^*(\theta))]\}dF(\theta) \\
&+ \int_{\theta_1^*}^{\theta_2^*}\{S-(1+\lambda)(rH(\theta)+(1-r)b_3) \\
&\quad +\alpha(1-r)[b_3-H(\theta-e^*(\theta))-\psi(e^*(\theta))]\}dF(\theta) \\
&+ \int_{\theta_2^*}^{\bar{\theta}}[S-(1+\lambda)H(\theta)]dF(\theta)
\end{aligned}$$

[2] Lemma (11) implies that $\theta_2^* = \theta_2^*(b_3)$ in $SW(C)$, and still $\theta_1^* = \theta_1^*(b_1, b_2, b_3)$.

The optimal renegotiation-proof menu solves the following optimization problem:[3]

$$\max_{C} SW(C) \quad \text{subject to (B.5).} \tag{$\mathbf{P}^R$}$$

Similar to the cut-off type θ^* under commitment, the two cut-off types under renegotiation satisfy

$$b_1 + \frac{(1-r)}{r}(b_2 - b_3) = H(\theta_1^* - e^*(\theta_1^*)) + \psi(e^*(\theta_1^*)), \tag{B.6}$$

$$b_3 = H(\theta_2^* - e(\theta_2^*)) + \psi(e(\theta_2^*)). \tag{B.7}$$

Due to the fact that $\frac{d\theta_1^*}{db_1} = \frac{r}{1-r}\frac{d\theta_1^*}{db_2}$ by (B.6), it is easy to show that the first-order conditions for b_1 and b_2 are the same, thus leading to the same optimal solution $b_1^R = b_2^R \equiv \underline{b}^R$. Denote the optimal solution for b_3 by $\bar{b}^R$. The first-order conditions with respect to b_1 and b_3 leads to

$$\frac{r(1+\lambda-\alpha)F(\theta_1^*)}{f(\theta_1^*)} = \frac{(1+\lambda)[rH(\theta_1^*) + (1-r)\bar{b}^R - \underline{b}^R] - \vartheta\left(1-\frac{\alpha}{1+\lambda}\right)}{H'(\theta_1^* - e^*(\theta_1^*, w))},$$

$$\left(1-\frac{\alpha}{1+\lambda}\right)[F(\theta_2^*) - F(\theta_1^*)] = \frac{[H(\theta_2^*) - \bar{b}^R]f(\theta_2^*)}{H'(\theta_2^* - e^*(\theta_2^*))} - \frac{[rH(\theta_1^*) + \bar{b}^R(1-r) - \underline{b}^R]f(\theta_1^*)}{rH'(\theta_1^* - e^*(\theta_1^*))} + \frac{\vartheta m(\theta_1^*, \theta_2^*, \lambda, r, \alpha)}{(1+\lambda)(1-r)},$$

where $\vartheta > 0$ is the Lagrange multiplier of the renegotiation-proof constraint (B.5), and

$$\begin{aligned} m(\theta_1^*, \theta_2^*, \lambda, r, \alpha) = {} & \left(1-\frac{\alpha}{1+\lambda}\right)\left(\frac{f(\theta_2^*)}{H'(\theta_2^* - e^*(\theta_2^*))} - \frac{f(\theta_1^*)(r-1)}{rH'(\theta_1^* - e^*(\theta_1^*))}\right) \\ & - \frac{[H'(\theta_2^*)f(\theta_2^*) + (H(\theta_2^*) - \bar{b}^R)f'(\theta_2^*)]H'[\theta_2^* - e^*(\theta_2^*)]}{[H'(\theta_2^* - e^*(\theta_2^*))]^3} \\ & + \frac{H''[\theta_2^* - e^*(\theta_2^*)][1 - e^{*\prime}(\theta_2^*)][H(\theta_2^*) - \bar{b}^R]f(\theta_2^*)}{[H'(\theta_2^* - e^*(\theta_2^*))]^3} \end{aligned}$$

[3] We assume (B.4) holds with strict inequality and (B.5) holds with equality as shown in Lemma 11.

This completes the proof of Proposition 2.

Proof of Lemma 3. Let $(\widetilde{C}, \widetilde{D}_1, \widetilde{D}_2, \widetilde{\overline{B}}^R, \widetilde{\underline{B}}^R)$ denote the endogenous variables under the structure $\widetilde{\mathbf{S}}$. In actuality, the equivalence between $\mathbf{S}$ and $\widetilde{\mathbf{S}}$ can be obtained by taking a linear transformation that $\widetilde{\theta} = \xi_1\theta$ without an intercept term, where $\xi_1 > 0$. To do this, let us first consider a general linear transformation that $\widetilde{\theta} = \xi_0 + \xi_1\theta$ with $(\xi_0, \xi_1) \in \mathbf{R}^2_+$, then the distribution of $\widetilde{\theta}$ is $\widetilde{F}(\cdot) = F((\cdot - \xi_0)/\xi_1)$. To justify the observational equivalence, we need to show that $(D_1, D_2, C, \overline{B}^R, \underline{B}^R) = (\widetilde{D}_1, \widetilde{D}_2, \widetilde{C}, \widetilde{\overline{B}}^R, \widetilde{\underline{B}}^R)$, and that the equality (3.10) holds under the structure $\widetilde{\mathbf{S}}$. Let $\widetilde{\theta}^*_1 = \xi_0 + \xi_1\theta^*_1$ and $\widetilde{\theta}^*_2 = \xi_0 + \xi_1\theta^*_2$, then for any $\widetilde{\theta}$, $\widetilde{\theta} \leq \widetilde{\theta}^*_1$ is equivalent to $\theta \leq \theta^*_1$, which implies that $\widetilde{D}_1 = D_1$. Similarly, we have $\widetilde{D}_2 = D_2$. Note that

$$\widetilde{\psi}'(\widetilde{e}^*) = \widetilde{H}'(\widetilde{\theta} - \widetilde{e}^*) \Rightarrow \psi'[(\widetilde{e}^* - \xi_0)/\xi_1] = H'[(\widetilde{\theta} - \widetilde{e}^*)/\xi_1],$$

which leads to $\widetilde{e}^*(\widetilde{\theta}) = \xi_0 + \xi_1 e^*(\theta)$. For those with FP contracts,

$$\begin{aligned}
\widetilde{C} &= \widetilde{H}(\widetilde{\theta} - \widetilde{e}^*(\widetilde{\theta}^*)) = H[(\xi_1\theta - \xi_1 e^*(\theta^*))/\xi_1] = H(\theta - e^*(\theta^*)) = C, \\
\widetilde{\overline{B}}^R &= \widetilde{H}(\widetilde{\theta}^*_2 - \widetilde{e}^*(\widetilde{\theta}^*_2)) + \widetilde{\psi}(\widetilde{e}^*(\widetilde{\theta}^*_2)) = H((\xi_1\theta^*_2 - \xi_1 e^*(\theta^*_2))/\xi_1) + \psi(\xi_1 e^*(\theta^*)/\xi_1) = \overline{B}^R, \\
\widetilde{\underline{B}}^R &= r[\widetilde{H}(\widetilde{\theta}^*_1 - \widetilde{e}^*(\widetilde{\theta}^*_1)) + \widetilde{\psi}(\widetilde{e}^*(\widetilde{\theta}^*_1))] + (1-r)\overline{B}^R \\
&= r[H(\theta^*_1 - e^*(\theta^*_1)) + \psi(e^*(\theta^*_1))] + (1-r)\overline{B}^R = \underline{B}^R.
\end{aligned}$$

For those associated with CR contracts, since $\widetilde{C} = \widetilde{H}(\widetilde{\theta}) = H(\widetilde{\theta}/\xi_1) = H((\xi_0 + \xi_1\theta)/\xi_1)$, then $\widetilde{C} = C$ is equivalent to $\xi_0 = 0$ by noting that $C = H(\theta)$. In what follows, we just need to consider that $\widetilde{\theta} = \xi_1\theta$. Since

$$\widetilde{f}(\widetilde{\theta}^*_j) = \frac{\partial \widetilde{F}(\widetilde{\theta}^*_j)}{\partial \widetilde{\theta}^*_j} = \frac{\partial F(\widetilde{\theta}^*_j/\xi_1)}{\partial \widetilde{\theta}^*_j} = f(\widetilde{\theta}^*_j/\xi_1)/\xi_1 = f(\theta^*_j)/\xi_1, \quad j = 1, 2$$

we have

$$\left(1-\frac{\alpha}{1+\lambda}\right)\frac{\widetilde{F}(\widetilde{\theta}_2^*)-\widetilde{F}(\widetilde{\theta}_1^*)}{\widetilde{f}(\widetilde{\theta}_2^*)}=\xi_1\left(1-\frac{\alpha}{1+\lambda}\right)\frac{F(\theta_2^*)-F(\theta_1^*)}{f(\theta_2^*)}=\xi_1\frac{H(\theta_2^*)-\overline{B}^R}{H'(\theta_2^*-e^*(\theta_2^*,w))}$$

and

$$\frac{\widetilde{H}(\widetilde{\theta}_2^*)-\widetilde{\overline{B}}^R}{\widetilde{H}'(\widetilde{\theta}_2^*-\widetilde{e}^*(\widetilde{\theta}_2^*))}=\frac{\widetilde{H}(\widetilde{\theta}_2^*)-\widetilde{\overline{B}}^R}{\widetilde{\psi}'(\widetilde{e}^*(\widetilde{\theta}))}=\frac{H(\theta_2^*)-\overline{B}^R}{\psi'(e^*)/\xi_1}=\xi_1\frac{H(\theta_2^*)-\overline{B}^R}{H'(\theta_2^*-e^*(\theta_2^*))}.$$

Hence,

$$\left(1-\frac{\alpha}{1+\lambda}\right)\frac{\widetilde{F}(\widetilde{\theta}_2^*)-\widetilde{F}(\widetilde{\theta}_1^*)}{\widetilde{f}(\widetilde{\theta}_2^*)}=\frac{\widetilde{H}(\widetilde{\theta}_2^*)-\widetilde{\overline{B}}^R}{\widetilde{H}'(\widetilde{\theta}_2^*-\widetilde{e}^*(\widetilde{\theta}_2^*))}.$$

This completes the proof.

Proof of Proposition 3. First note that the strictly increasing property of $H(\cdot)$ leads to $\theta(\tau) = e_j(\tau) + H^{-1}(C_j(\tau))$ for any $\tau \in [0,1]$, where $\theta(\tau)$ denotes the $\tau-$quantile of $F(\cdot)$, $e_j(\tau)$ denotes the $\tau-$quantile of $F_E(\cdot, W = w_j)$, and $C_j(\tau)$ denotes the $\tau-$quantile of $F_{C_j}(\cdot)$. Hence, we obtain the compatibility condition

$$e_1(\tau)+H^{-1}(C_1(\tau))=e_2(\tau)+H^{-1}(C_2(\tau)), \tag{B.8}$$

that is,

$$H^{-1}(C_1(\tau))=H^{-1}(C_2(\tau))+e_2(\tau)-e_1(\tau)=H^{-1}(C_2(\tau))+\Delta e(\tau), \tag{B.9}$$

where $\Delta e(\tau) \equiv e_2(\tau) - e_1(\tau)$. Taking the first derivative with respect to W in both sides of (3.4) and adding W back implies

$$\psi_{11}(e^*(\theta,W),W)e_2^*(\theta,W)+\psi_{12}(e^*(\theta,W),W)=-H''(\theta-e^*(\theta,W))e_2^*(\theta,W),$$

which implies that $e_2^*(\theta,W)>0$ for $\theta\in(\underline{\theta},\theta^*]$ due to the assumption $\psi_{12}(e^*(\theta),W)<0$, and that $e_2^*(\underline{\theta},W)=0$ due to the normalization assumption $\psi_{12}(e^*(\underline{\theta}),W)=0$. The latter result further implies that $C_1(0)=C_2(0)\equiv\underline{c}$ with the corresponding type $\underline{\theta}$ for $\underline{c}$ due to $\Delta e(0)=0$ and $\theta(0)=\underline{\theta}$. Therefore, $\underline{e}\equiv e_1(\theta_0,w_1)=e_2(\theta_0,w_2)$, and for $\tau\in(0,1]$, $\Delta e(\tau)>0$, $H^{-1}(C_1(\tau))>H^{-1}(C_2(\tau))$, and in turn we obtain that $C_1(\tau)>C_2(\tau)$, and this proves Lemma 5.

Let some $x\in[\underline{c},c_{H,1}]$, where $c_{H,1}\equiv H(\theta_2^*-e^*(\theta_2^*,w_1))$ due to the fact that $\partial C/\partial\theta=H'(\theta-e^*(\theta,W))(1-e_1^*(\theta,W))>0$ and the assumption that $w_1<w_2$. If $x=\underline{c}$, then $H^{-1}(\underline{c})=\underline{\theta}-\underline{e}$. If $x\in(\underline{c},c_{H,1}]$, based on the facts that $C=H(\theta-e^*(\theta,W))$ is strictly increasing in θ and that $C_1(\cdot)$ is continuous on $[0,1]$, there exists a unique $\tau_0\in(0,1]$ such that $C_1(\tau_0)=x$. In particular, we obtain $x=C_1(\tau_0)>C_2(\tau_0)>\underline{c}$. Similarly, there exists a unique $\tau_1\in(0,\tau_0)$ such that $C_1(\tau_1)=C_2(\tau_0)$. Continuing such a construction gives rise to $C_1(\tau_1)>C_2(\tau_1)>\underline{c}$, which in turn implies that there exists a unique $\tau_2\in(0,\tau_1)$ such that $C_1(\tau_2)=C_2(\tau_1)$. Thereby, we have constructed a unique sequence such that $1\geq\tau_0>\tau_1>\cdots>\tau_t>\cdots>0$ with $x=C_1(\tau_0)>C_2(\tau_0)=C_1(\tau_1)>C_2(\tau_1)=C_1(\tau_2)>\cdots>C_2(\tau_{t-1})=C_1(\tau_t)>\cdots>\underline{c}$. Since the sequence $\{\tau_t\}$ is strictly decreasing, $\{\tau_t\}$ converges to some $\underline{\tau}\in[0,1]$ as $t\to\infty$. Also, the continuity of $C_j(\cdot)$ on $[0,1]$ implies that $\lim_{t\to\infty}C_j(\tau_t)=C_j(\underline{\tau})$. Note that $C_2(\tau_t)=C_1(\tau_{t+1})$ implies $C_2(\underline{\tau})=C_1(\underline{\tau})$ and that $C_2(\tau)=C_1(\tau)$ only for $\tau=0$, we obtain $\underline{\tau}=0$. As a result, $\lim_{t\to\infty}H^{-1}(C_j(\tau_t))=H^{-1}(\underline{\tau})=\underline{\theta}$ for $j=1,2$ by the finiteness of the support of θ and the continuity of $H(\cdot)$. By iterating (B.9) we obtain the following nonlinear dynamic relation

$$\begin{aligned}
H^{-1}(x) &= H^{-1}(C_2(\tau_0)) + \Delta e(\tau_0) = H^{-1}(C_1(\tau_1)) + \Delta e(\tau_0) \\
&= H^{-1}(C_2(\tau_1)) + \Delta e(\tau_0) + \Delta e(\tau_1) \\
&= \cdots = H^{-1}(C_2(\tau_t)) + \Delta e(\tau_0) + \cdots + \Delta e(\tau_t).
\end{aligned}$$

Because $H^{-1}(x)$ is finite, it follows that $\sum_{t=0}^{\infty} \Delta e(\tau_t)$ must exist, thus leading to $H^{-1}(x) = \underline{\theta} - \underline{e} + \sum_{t=0}^{\infty} \Delta e(\tau_t) = \theta_0 - \underline{e} + \sum_{t=0}^{\infty} \Delta e(\tau_t)$ by using the normalization $\underline{\theta} = \theta_0$, i.e., $H^{-1}(\cdot)$ and hence $H(\cdot)$ are identified on $[\underline{c}, c_{H,1}]$ from the observed costs associated with fixed-price contracts.

As stated in the text, the identification of $H(\cdot)$ is valid for the discrete W. Specifically, the corresponding modification of Assumption 4 is that in (ii) $\psi_1(e^*(\underline{\theta}, w_1), w_1) = \psi_1(e^*(\underline{\theta}, w_2), w_2)$ for any $(w_1, w_2) \in \mathbf{W}^2$, and that in (iii) if $w_2 > w_1$, then $\psi_1(e^*(\theta, w_1), w_1) > \psi_1(e^*(\theta, w_2), w_2)$ for any $\theta \in (\underline{\theta}, \theta_2^*]$. Under this modified assumption, the relevant modification of the above proof is as follows. Using (3.4) leads to

$$\psi_1(e^*(\theta, w_2), w_2) - \psi_1(e^*(\theta, w_1), w_1) = H'(\theta - e^*(\theta, w_2)) - H'(\theta - e^*(\theta, w_1)),$$

which implies that $e^*(\theta, w_2) > e^*(\theta, w_1)$ for $\theta \in (\underline{\theta}, \theta^*]$ and that $e^*(\underline{\theta}, w_2) = e^*(\underline{\theta}, w_1)$. The remaining arguments hold without any modification.

B.2 Further Details for Empirical Application

We first provide some details of parametric identification. In the specification of cost function (3.36), the parameters $\widetilde{\gamma}_0, \widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$ can be expressed in terms of $\beta_1, \beta_2, \gamma_1, \gamma_2$ and

σ_η^2 as follows.

$$\begin{aligned}
\widetilde{\gamma}_0 &= (\beta_2\gamma_1^2 - \beta_1^2\beta_2 - 2\beta_1^2\gamma_2 + 2\gamma_1\gamma_2\beta_1 + 4\beta_2\gamma_2^2\sigma_\eta^2)[2(\beta_2+\gamma_2)]^{-2}, \\
\widetilde{\gamma}_1 &= (\beta_1\gamma_2^2 + \gamma_1\gamma_2\beta_2)(\beta_2+\gamma_2)^{-2}, \\
\widetilde{\gamma}_2 &= \beta_2\gamma_2^2(\beta_2+\gamma_2)^{-2}, \\
\varepsilon_2 &= 2\beta_2\gamma_2^2(\beta_2+\gamma_2)^{-2}x\eta + (\beta_1\gamma_2^2 + \gamma_1\gamma_2\beta_2)(\beta_2+\gamma_2)^{-2}\eta,
\end{aligned}$$

where the composite error ε_2 satisfies $E[\varepsilon_2|x,z] = 0$. The parameters $\widetilde{\gamma}_0, \widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$ are directly estimated from data, then the equations above allows us to recover γ_2 and γ_1 under Assumption 2:

$$\begin{aligned}
\gamma_1 &= (\beta_2\gamma_2)^{-1}[\widetilde{\gamma}_1(\beta_2+\gamma_2)^2 - \beta_1\gamma_2^2], \\
\gamma_2 &= (\beta_2 - \widetilde{\gamma}_2)^{-1}(\beta_2\widetilde{\gamma}_2 + \beta_2\sqrt{\beta_2\widetilde{\gamma}_2}).
\end{aligned}$$

By now, all the primitive parameters of the cost and disutility functions $(\beta_1, \beta_2, \beta_3, \widetilde{\beta}_3, \gamma_1, \gamma_2)$ are identified.

Next, we present the parametric form of $\rho_0(\cdot)$ and $\rho_1(\cdot)$ in (3.37)

$$\begin{aligned}
\rho_0(z) &= (\beta_2+\gamma_2)^{-1}[\beta_2\gamma_2\theta_2^*(\widetilde{z})^2 + (\beta_1\gamma_2 + \beta_2\gamma_1)\theta_2^*(\widetilde{z}) - (\beta_1-\gamma_1)^2/4], \\
\rho_1(z) &= (\beta_2+\gamma_2)^{-1}\left\{\beta_2\gamma_2[r\theta_1^*(z)^2 + (1-r)\theta_2^*(z)^2]\right. \\
&\quad \left. +(\beta_1\gamma_2 + \beta_2\gamma_1)[r\theta_1^*(z) + (1-r)\theta_2^*(z)] - (\beta_1-\gamma_1)^2/4\right\}.
\end{aligned}$$

Lastly, we give the definition of welfare related terms. $T^R(z)$ and $T^C(z)$ are subsidy (tax) under renegotiation and commitment, respectively, defined as follows.

$$\begin{aligned}
T^R(z) &= \underline{b}^R F(\theta_1^*(z); \mu(z), \sigma^2) + \int_{\theta_1^*(z)}^{\theta_2^*(z)} [rH(\theta, z) + (1-r)\bar{b}^R] dF(\theta; \mu(z), \sigma^2) \\
&\quad + \int_{\theta_2^*(z)}^{\infty} H(\theta, z) dF(\theta; \mu(z), \sigma^2), \\
T^C(z) &= b^C F(\theta^*(z); \mu(z), \sigma^2) + \int_{\theta^*(z)}^{\infty} H(\theta, z) dF(\theta; \mu(z), \sigma^2).
\end{aligned}$$

$U^R(z)$ and $U^C(z)$ are the informational rent (profit) counterparts.

$$\begin{aligned}
U^R(z) &= \int_{-\infty}^{\theta_1^*(z)} [\underline{b}^R - H(\theta - e(\theta), z) - \psi(e(\theta))] dF(\theta; \mu(z), \sigma^2) \\
&+ (1-r) \int_{\theta_1^*(z)}^{\theta_2^*(z)} [\bar{b}^R - H(\theta - e(\theta), z) - \psi(e(\theta))] dF(\theta; \mu(z), \sigma^2) \\
U^C(z) &= \int_{-\infty}^{\theta^*(z)} [b^C - H(\theta - e^*(\theta), z) - \psi(e(\theta))] dF(\theta; \mu(z), \sigma^2).
\end{aligned}$$

APPENDIX C

NONPARAMETRIC IDENTIFICATION AND ESTIMATION OF CONTRACTING WITH EXTERNALITIES

C.1 Proof of Lemma 8

We complete the proof by verifying the assumptions in Theorem 3.3 of [40]. First, for each $z \in \mathcal{Z}$, the function $\boldsymbol{m}(\boldsymbol{q}, z)$ is a one-to-one correspondence because for each $(\boldsymbol{\epsilon}, z)$, there is a unique $\boldsymbol{q}$ due to the relationship in (4.2) and $s'(\cdot, z) > 0$. Second, for each $(\boldsymbol{q}, z)$, the Jacobian determinant det$(\partial\widetilde{\boldsymbol{m}}(\boldsymbol{q}, z)/\partial\boldsymbol{q}) > 0$. Specifically,

$$\frac{\partial\widetilde{m}^i(\boldsymbol{q}, z)}{\partial q_i} = \frac{1}{N}s'(x, z) + \frac{1-N}{N}\widetilde{u}'(x, z) + \frac{x}{N}\widetilde{u}''(x, z) - q_i\widetilde{u}''(x, z) \equiv \widetilde{b}_1(x, z) - q_i\widetilde{u}''(x, z),$$

and for any $j \neq i$,

$$\frac{\partial\widetilde{m}^i(\boldsymbol{q}, z)}{\partial q_j} = \frac{1}{N}s'(x, z) + \frac{1}{N}\widetilde{u}'(x, z) + \frac{x}{N}\widetilde{u}''(x, z) - q_i\widetilde{u}''(x, z) \equiv \widetilde{b}_2(x, z) - q_i\widetilde{u}''(x, z),$$

where $\widetilde{b}_2(x, z) = \widetilde{b}_1(x, z) + \widetilde{u}'(x, z)$. It follows
det$(\partial\widetilde{\boldsymbol{m}}(\boldsymbol{q}, z)/\partial\boldsymbol{q}) = [\widetilde{u}'(x, z)]^{N-1}s'(x, z) > 0$
due to $\widetilde{u}'(\cdot) > 0$ and $s'(\cdot) > 0$. Last, the definitions of Γ and $\mathcal{F}_\epsilon$ guarantee that $\widetilde{\boldsymbol{m}}$ is twice continuously differentiable on its support. The proof is complete.

C.2 Proof of Proposition 4

Since

$$\begin{aligned}
\log\left(\det\left(\frac{\partial\boldsymbol{m}(\mathbf{q}, N, z)}{\partial\boldsymbol{q}}\right)\right) &= (N-1)\log(u'(x, z)) + \log(s'(x, z)) \\
\log\left(\det\left(\frac{\partial\widetilde{\boldsymbol{m}}(\boldsymbol{q}, N, z)}{\partial\boldsymbol{q}}\right)\right) &= (N-1)\log(\widetilde{u}'(x, z)) + \log(s'(x, z))
\end{aligned}$$

then

$$\begin{aligned}
&\Delta_{\boldsymbol{q}}(\boldsymbol{q}, z; \partial \boldsymbol{m}, \partial^2 \boldsymbol{m}, \partial \widetilde{\boldsymbol{m}}, \partial^2 \widetilde{\boldsymbol{m}}) \\
&= \left(\frac{(N-1)u''(x,z)}{u'(x,z)} - \frac{(N-1)\widetilde{u}''(x,z)}{\widetilde{u}'(x,z)}, \cdots, \frac{(N-1)u''(x,z)}{u'(x,z)} - \frac{(N-1)\widetilde{u}''(x,z)}{\widetilde{u}'(x,z)} \right)',
\end{aligned}$$

and

$$\Delta_z(\boldsymbol{q}, z; \partial \boldsymbol{m}, \partial^2 \boldsymbol{m}, \partial \widetilde{\boldsymbol{m}}, \partial^2 \widetilde{\boldsymbol{m}}) \;=\; \frac{N-1}{u'(x,z)} \times \frac{\partial u'(x,z)}{\partial z} - \frac{N-1}{\widetilde{u}'(x,z)} \times \frac{\partial \widetilde{u}'(x,z)}{\partial z} = 0.$$

Let

$$\begin{aligned}
\widetilde{b}_1 &\equiv \widetilde{b}_1(x,z) = N^{-1}s'(x,z) + (N^{-1}-1)\widetilde{u}'(x,z) + N^{-1}x\widetilde{u}''(x,z), \\
\widetilde{b}_2 &\equiv \widetilde{b}_2(x,z) = \widetilde{b}_1(x,z) + \widetilde{u}'(x,z), \\
\widetilde{b}_3 &\equiv \widetilde{b}_3(x,z) = N^{-1}\partial \widetilde{s}(x,z)/\partial z,
\end{aligned}$$

Let $z = z^0(\boldsymbol{q})$, under Assumption 6 (i)-(ii), (4.3) reduces to

$$\begin{pmatrix}
\widetilde{b}_1 - q_1\widetilde{u}'' & \cdots & \widetilde{b}_2 - q_N\widetilde{u}'' & \frac{(N-1)u''}{u'} - \frac{(N-1)\widetilde{u}''}{\widetilde{u}'} \\
\widetilde{b}_2 - q_1\widetilde{u}'' & \cdots & \widetilde{b}_2 - q_N\widetilde{u}'' & \frac{(N-1)u''}{u'} - \frac{(N-1)\widetilde{u}''}{\widetilde{u}'} \\
\cdots & \cdots & \cdots & \cdots \\
\widetilde{b}_2 - q_1\widetilde{u}'' & \cdots & \widetilde{b}_1 - q_N\widetilde{u}'' & \frac{(N-1)u''}{u'} - \frac{(N-1)\widetilde{u}''}{\widetilde{u}'} \\
\widetilde{b}_3 & \cdots & \widetilde{b}_3 & 0
\end{pmatrix} \tag{C.1}$$

Note that the ranks of both matrix (C.1) and matrix $\left(\frac{\partial \widetilde{\boldsymbol{m}}(\boldsymbol{q},N,z)}{\partial \boldsymbol{q}}\right)$ are N due to det$(\partial \widetilde{\boldsymbol{m}}(\boldsymbol{q}, z)/\partial \boldsymbol{q}) > 0$. According to Assumption 6 (iii), the zero element in the right bottom of the matrix (C.1) implies that there exists one zero coefficient such that the last column is a linear combination of the other N columns. Hence,

$$\frac{u''}{u'} = \frac{\widetilde{u}''}{\widetilde{u}'}.$$

Since u_1''/u_1' does not depend on z, u_1''/u_1' is identified. Now we can identify $u_1'(x)$ by

$$u_1'(x) = \underline{u}_1' exp\left(\int_{\underline{x}}^{x} \frac{u_1''(t)}{u_1'(t)} dt\right),$$

and then identify $u_1(x)$ through

$$u_1(x) = \int_{\underline{x}}^{x} u_1'(t) + \underline{u}_1.$$

The proof is complete.

C.3 Proof of Proposition 5

Under Assumption 7-(i), it follows $s(x,z) = s_1(x) - Nu_2(z)$, where $s_1(x) = Nc'(x) - Nu_1(x) - xu_1'(x)$. According to [40], the ratio of first derivatives $s_2(x,z) \equiv s_1'(x)/Nu_2'(z)$ can be identified. Under normalization (ii), $s_1'(x) = s_2(x,z)s_1'(\underline{x})/s_2(\underline{x},z)$, where $s_1'(\underline{x}) = N\underline{c}'' - (N+1)\underline{u}_1' - \underline{x}u_1''(\underline{x})$. Hence, $u_2'(z) = N^{-1}s_2(x,z)/s_1(x)'$. Then, $u_2(z) = \int_{\underline{z}}^{z} u_2'(t)dt + \underline{u}_2$, and $c''(x) = N^{-1}[s_2(x,z)Nu_2'(z) + (N+1)u_1'(x) + xu_1''(x)]$. Then $c'(x)$ can be identified as $c'(x) = \int_{\underline{x}}^{x} c''(t)dt + \underline{c}'$, and hence $c(x) = \int_{\underline{x}}^{x} c'(t)dt + \underline{c}$. Now we can back out the pseudo ϵ_i for each $i \in N$ by using the optimal condition (4.1) and identify the distribution $F_{\boldsymbol{\epsilon}}(\cdot)$.

C.4 Proof of Lemma 9

Assumption 3.1(i) is satisfied under homoskedasticity, Assumption 3.1(ii) is not necessary given the identification result in Section 3, Assumption 3.1(iii) is satisfied because

$$\begin{aligned} ||\varpi - \Pi_T\varpi||_s &= \sup_{(x,z)}|[c(x,z) - \Pi_T c(x,z)]| \\ &+ \sup_{(x,z)}|[u'(x,z) - \Pi_T u'(x,z)]| + \sup_{z}|[\gamma(z) - \Pi_T\gamma(z)]| \\ &\leq c_1(K_{1T})^{-\iota/(1+d_z)} + c_2(K_{1T})^{-\iota/(1+d_z)} + c_3(K_{1T})^{-\iota/d_z} = o(1), \end{aligned}$$

where the last equality is obtained from Assumption 12-(i). Assumption 3.2 is trivially satisfied for the sieve minimum method. Under more primitive Assumptions 9 and 10, Assumption 3.3 holds for our series LS estimator $\widehat{m}(V,\varpi)$ by Lemma C.2 in [39]. Assumption 12-(ii) guarantees that $\mathcal{F}$ is compact under the sup norm, which implies that (12) in [39] is trivially satisfied. The proof is complete.

C.5 Proof of Proposition 8

Given $\boldsymbol{e} = (e_1, \cdots, e_N) \in \mathcal{S}_\epsilon$, define the kernel estimator of $f_{\boldsymbol{\epsilon}}$ with the true sample as $\widetilde{f}_{\boldsymbol{\epsilon}}(\boldsymbol{e}) = (Th_\epsilon^N)^{-1}\sum_{t=1}^T\prod_{i=1}^N k_\epsilon\left(\frac{e_i-\epsilon_{t,i}}{h_\epsilon}\right)$,

$$\begin{aligned} & \underset{\boldsymbol{e}\in\mathcal{S}_\epsilon}{sup}\left|\widehat{f}_{\boldsymbol{\epsilon}}(\boldsymbol{e}) - f_{\boldsymbol{\epsilon}}(\boldsymbol{e})\right| \leq \underset{\boldsymbol{e}\in\mathcal{S}_\epsilon}{sup}\left|\widehat{f}_{\boldsymbol{\epsilon}}(\boldsymbol{e}) - \widetilde{f}_{\boldsymbol{\epsilon}}(\boldsymbol{e})\right| + \underset{\boldsymbol{e}\in\mathcal{S}_\epsilon}{sup}\left|\widetilde{f}_{\boldsymbol{\epsilon}}(\boldsymbol{e}) - f_{\boldsymbol{\epsilon}}(\boldsymbol{e})\right| \\ = \; & \underset{\boldsymbol{e}\in\mathcal{S}_\epsilon}{sup}\left|T^{-1}\sum_{t=1}^T\sum_{i=1}^N\left[\left(\widehat{\epsilon}_{t,i} - \epsilon_{t,i}\right)h_\epsilon^{-N}k'_\epsilon\left(\frac{\widetilde{e}_i - \epsilon_{t,i}}{h_i}\right)\prod_{j\neq i}^N k_\epsilon\left(\frac{\widetilde{e}_j - \epsilon_{t,j}}{h_j}\right)\right]\right| \\ & + O_p\left\{\left[(Th_\epsilon^N)^{-1}\ln T\right]^{1/2} + h_\epsilon^2\right\} \\ = \; & O_p\left\{(a_T + b_T)h_\epsilon^{-N} + \left[(Th_\epsilon^N)^{-1}\ln T\right]^{1/2} + h_\epsilon^2\right\} = o_p(1). \end{aligned}$$

where the second term in the first equality follows [92] who build on [93].